



Software

Bagging

Legal Notices and Disclaimers

This presentation is for informational purposes only. INTEL MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at [intel.com](https://www.intel.com).

This sample source code is released under the [Intel Sample Source Code License Agreement](#).

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

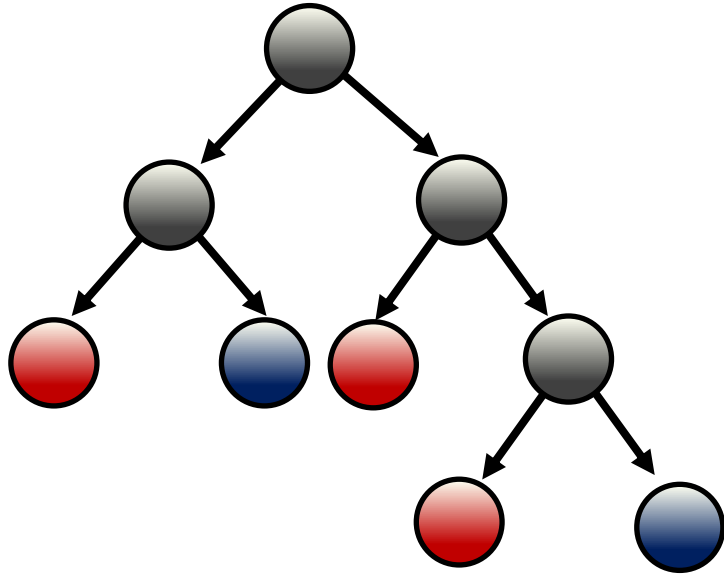
*Other names and brands may be claimed as the property of others.

Copyright © 2021, Intel Corporation. All rights reserved.

Learning Objectives

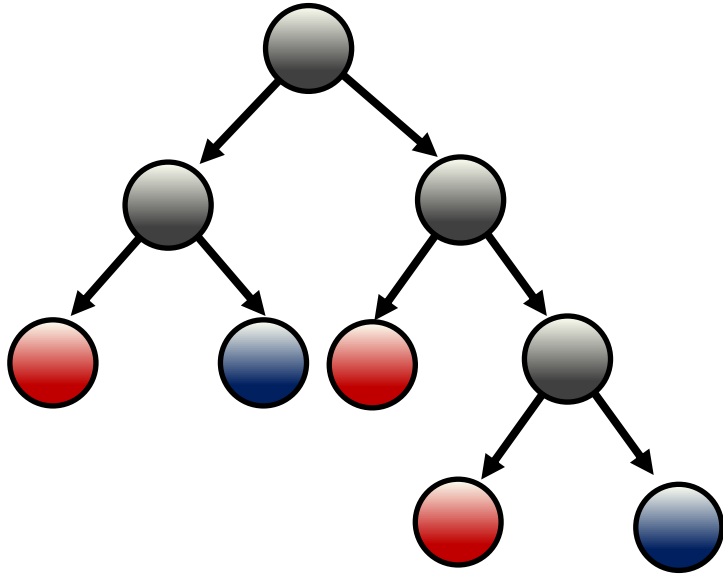
- Associate concepts of bootstrapping and aggregating (commonly known as “bagging”) to reduce variance
- Apply Random Forest algorithm that further reduces the correlation seen in bagging models
- Apply Intel® Extension for Scikit-learn* to leverage underlying compute capabilities of hardware

Decision Trees are High Variance



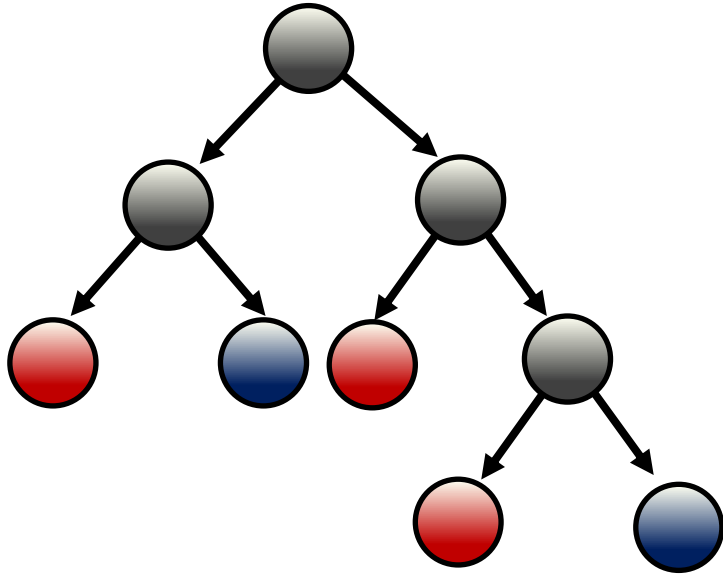
- Problem: decision trees tend to overfit

Decision Trees are High Variance



- Problem: decision trees tend to overfit
- Pruning helps reduce variance to a point

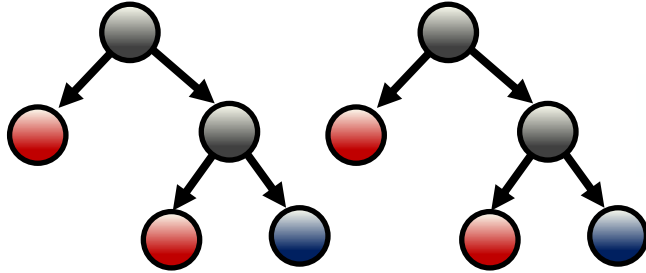
Decision Trees are High Variance



- Problem: decision trees tend to overfit
- Pruning helps reduce variance to a point
- Often not significant for model to generalize well

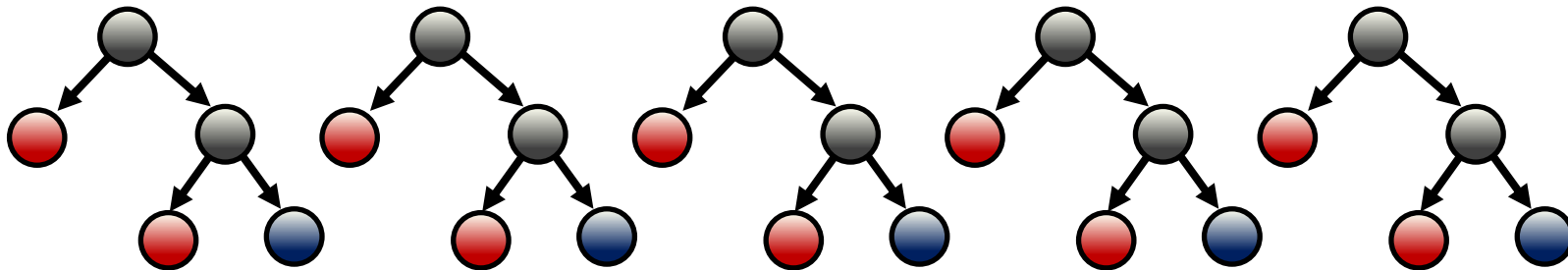
Improvement: Use Many Trees

Create many different trees



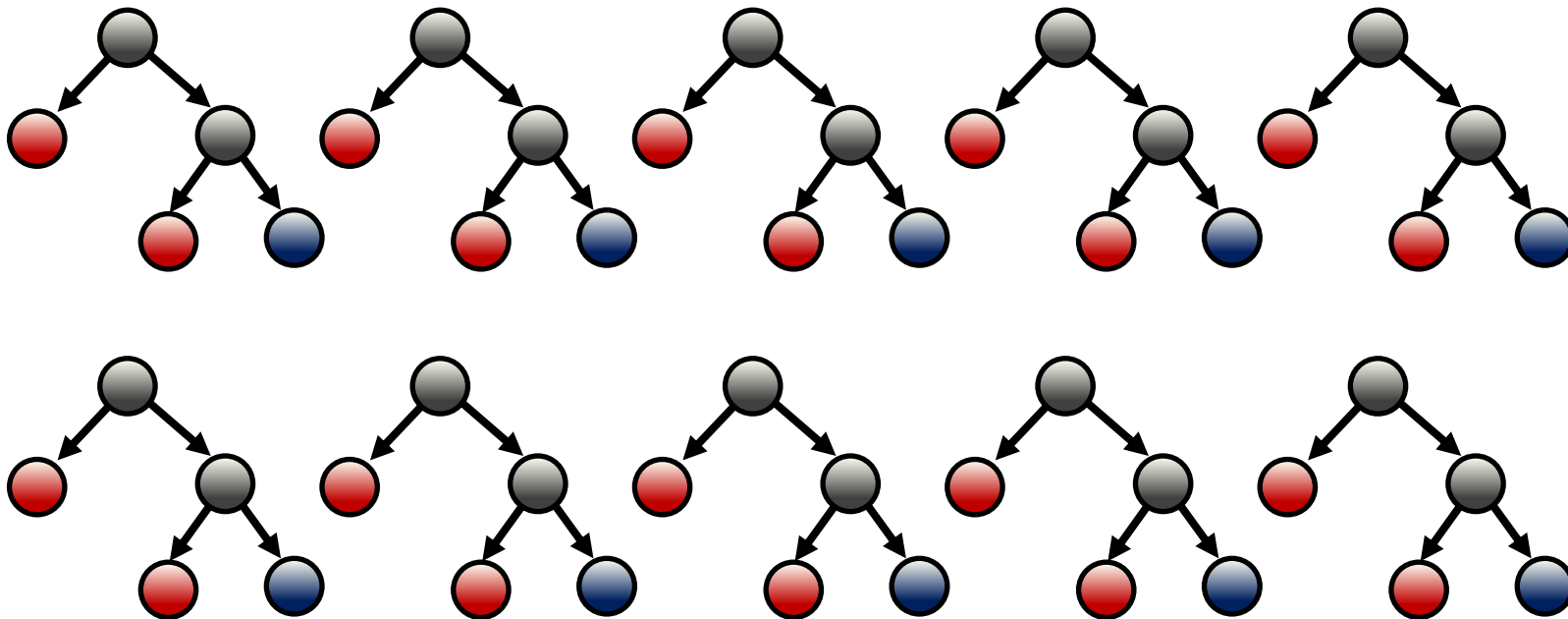
Improvement: Use Many Trees

Create many different trees



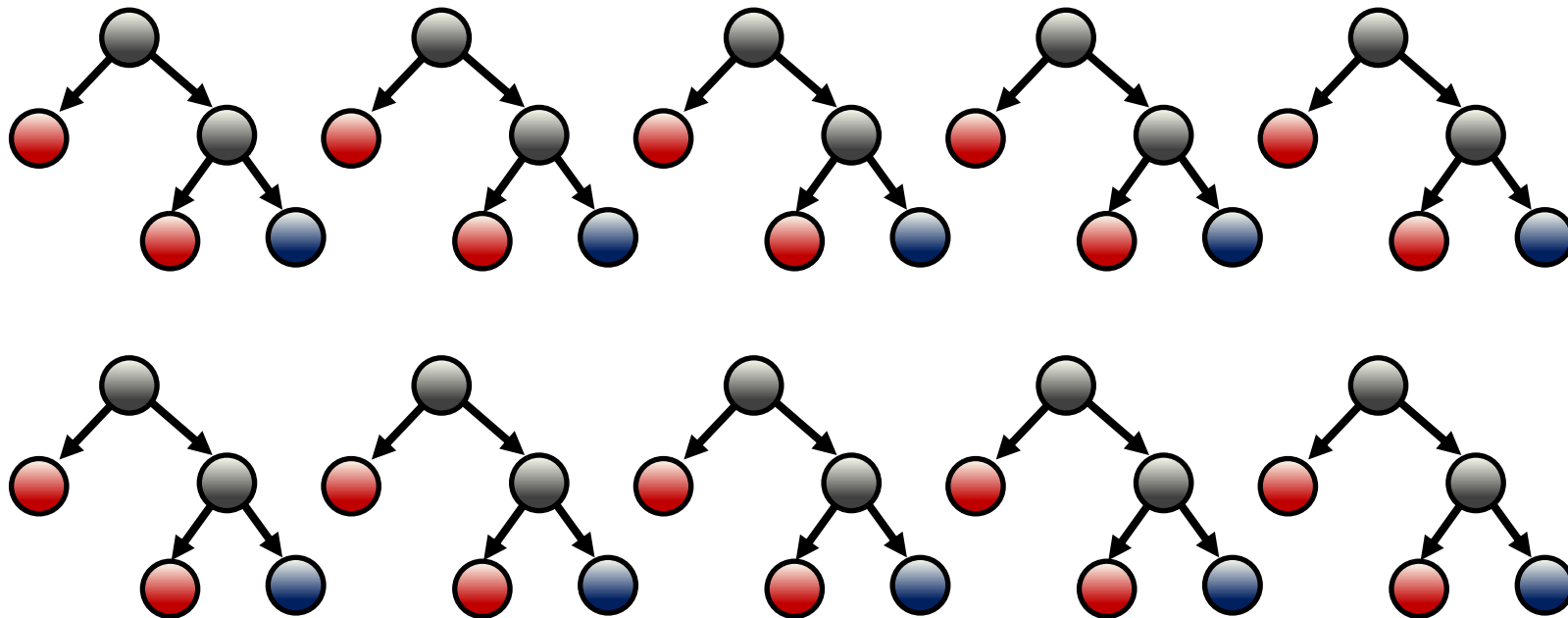
Improvement: Use Many Trees

Create many different trees



Improvement: Use Many Trees

Combine predictions to reduce variance



How to Create Multiple Trees?

Use bootstrapping: sample data with replacement

	Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22	The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03	Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22	Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03	Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14	Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04	Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-01	Monsters University	NaN	268482764	Dan Scanlon	G	107
7	2013-12-13	The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24	Fast & Furious 6	160000000	239679850	Justin Lin	PG-13	130
9	2013-03-08	Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16	Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08	Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21	World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22	The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28	The Heat	43000000	159562188	Paul Fieg	R	117
15	2013-08-07	We're the Millers	37000000	150294119	Rawson Marshall Thurber	R	110
16	2013-12-13	American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10	The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

How to Create Multiple Trees?

Create multiple bootstrapped samples

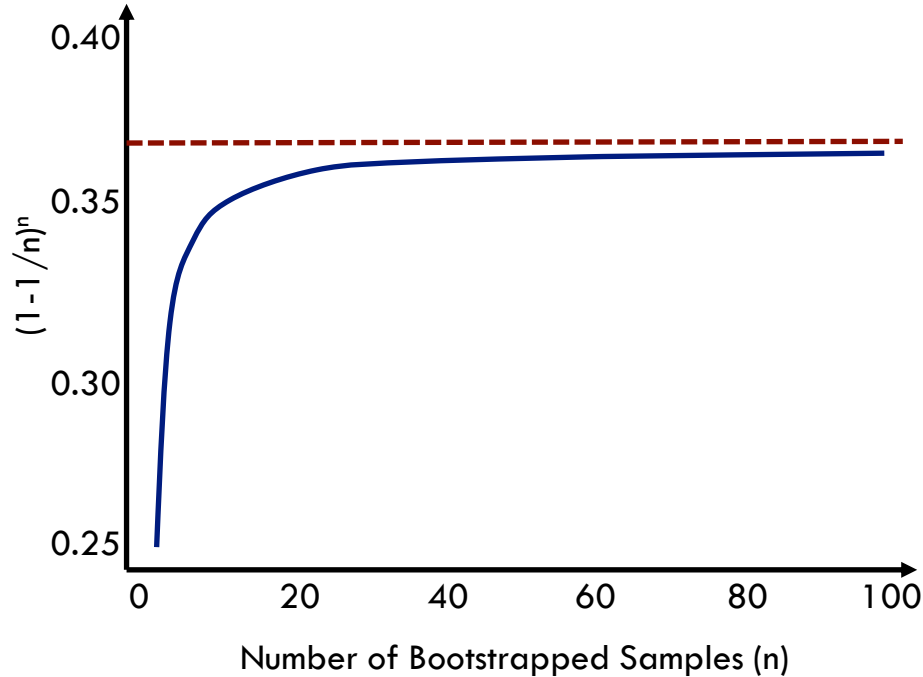
Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	13000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	20000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	15000000	400738009	Chris Buck/Jennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	7800000	368061265	Pierre Coffin/Chris Renaud	PG	98
4	2013-06-14 Man of Steel	22500000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	10000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NAH	268482764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NAH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	16000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	21500000	234911825	Sam Rami	PG	127
10	2013-05-16 Star Trek Into Darkness	19000000	228977861	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	17000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	19000000	202359711	Marc Forster	PG-13	116
13	2013-03-02 The Croods	13000000	187168425	Kirk De Micco/Chris Sanders	PG	98
14	2013-06-28 The Heat	4300000	159562188	Paul Fieg	R	117
15	2013-08-07 We're the Millers	3700000	150394119	Roxann Marshall/Thurber	R	110
16	2013-12-13 American Hustle	4000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	10500000	144840419	Baz Luhrmann	PG-13	143

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	13000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	20000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	15000000	400738009	Chris Buck/Jennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	7800000	368061265	Pierre Coffin/Chris Renaud	PG	98
4	2013-06-14 Man of Steel	22500000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	10000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NAH	268482764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NAH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	16000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	21000000	234911825	Sam Rami	PG	127
10	2013-05-16 Star Trek Into Darkness	19000000	228778861	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	17000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	19000000	202359711	Marc Forster	PG-13	116
13	2013-03-02 The Croods	13000000	187168425	Kirk De Micco/Chris Sanders	PG	98
14	2013-06-28 The Heat	4300000	159562188	Paul Fieg	R	117
15	2013-08-07 We're the Millers	3700000	150394119	Roxann Marshall/Thurber	R	110
16	2013-12-13 American Hustle	4000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	10500000	144840419	Baz Luhrmann	PG-13	143

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	13000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	20000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	15000000	400738009	Chris Buck/Jennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	7800000	368061265	Pierre Coffin/Chris Renaud	PG	98
4	2013-06-14 Man of Steel	22500000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	10000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NAH	268482764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NAH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	16000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	21000000	234911825	Sam Rami	PG	127
10	2013-05-16 Star Trek Into Darkness	19000000	228778861	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	17000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	19000000	202359711	Marc Forster	PG-13	116
13	2013-03-02 The Croods	13000000	187168425	Kirk De Micco/Chris Sanders	PG	98
14	2013-06-28 The Heat	4300000	159562188	Paul Fieg	R	117
15	2013-08-07 We're the Millers	3700000	150394119	Roxann Marshall/Thurber	R	110
16	2013-12-13 American Hustle	4000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	10500000	144840419	Baz Luhrmann	PG-13	143

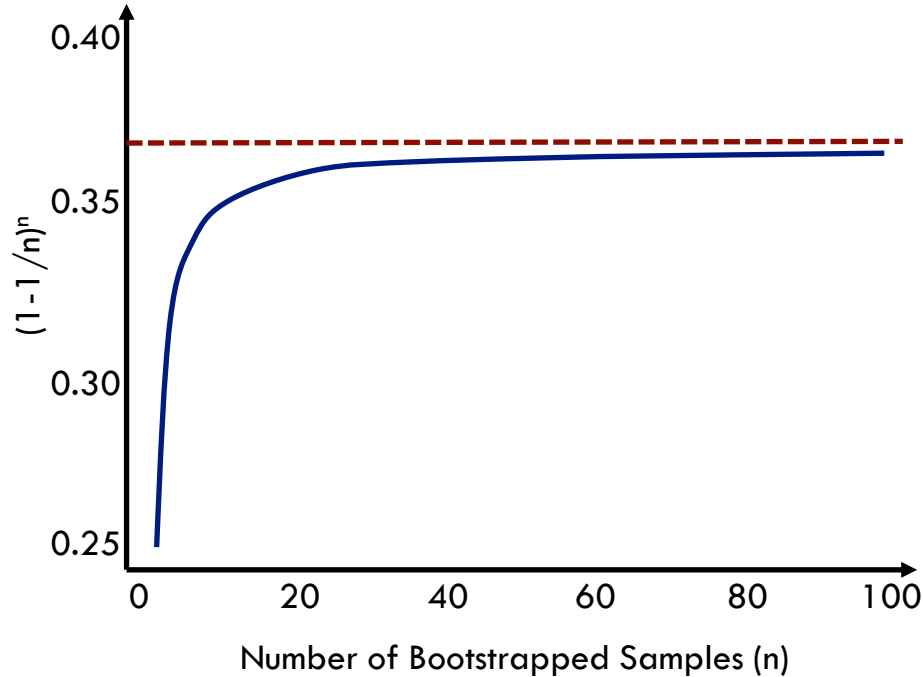
Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	13000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	20000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	15000000	400738009	Chris Buck/Jennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	7800000	368061265	Pierre Coffin/Chris Renaud	PG	98
4	2013-06-14 Man of Steel	22500000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	10000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NAH	268482764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NAH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	16000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	21000000	234911825	Sam Rami	PG	127
10	2013-05-16 Star Trek Into Darkness	19000000	228778861	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	17000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	19000000	202359711	Marc Forster	PG-13	116
13	2013-03-02 The Croods	13000000	187168425	Kirk De Micco/Chris Sanders	PG	98
14	2013-06-28 The Heat	4300000	159562188	Paul Fieg	R	117
15	2013-08-07 We're the Millers	3700000	150394119	Roxann Marshall/Thurber	R	110
16	2013-12-13 American Hustle	4000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	10500000	144840419	Baz Luhrmann	PG-13	143

Distribution of Data in Bootstrapped Samples



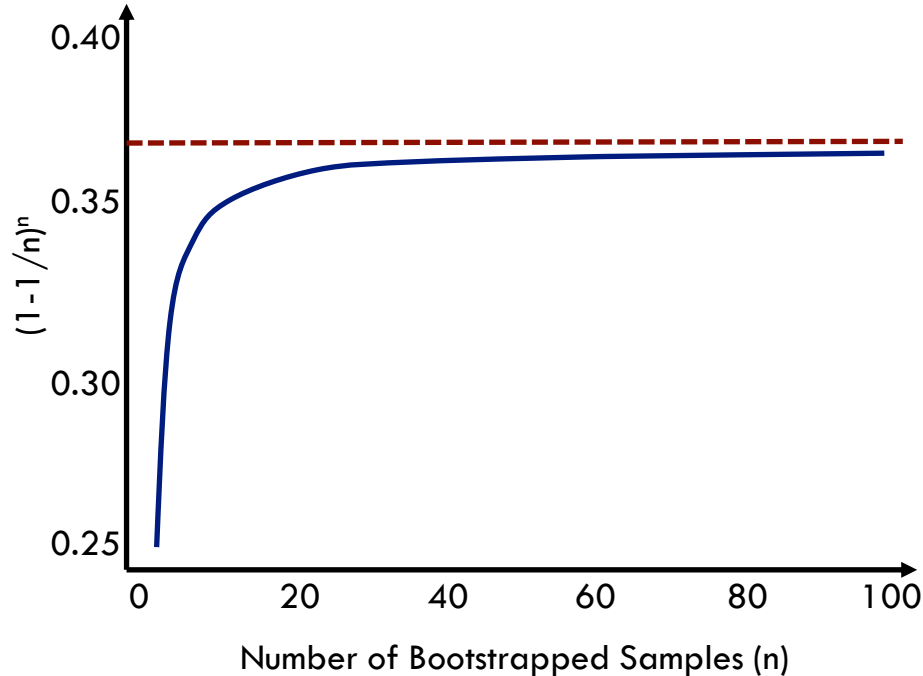
- Given a dataset, create n bootstrapped samples

Distribution of Data in Bootstrapped Samples



- Given a dataset, create n bootstrapped samples
- For a given record x ,
 $P(\text{rec } x \text{ not selected}) = (1 - 1/n)^n$

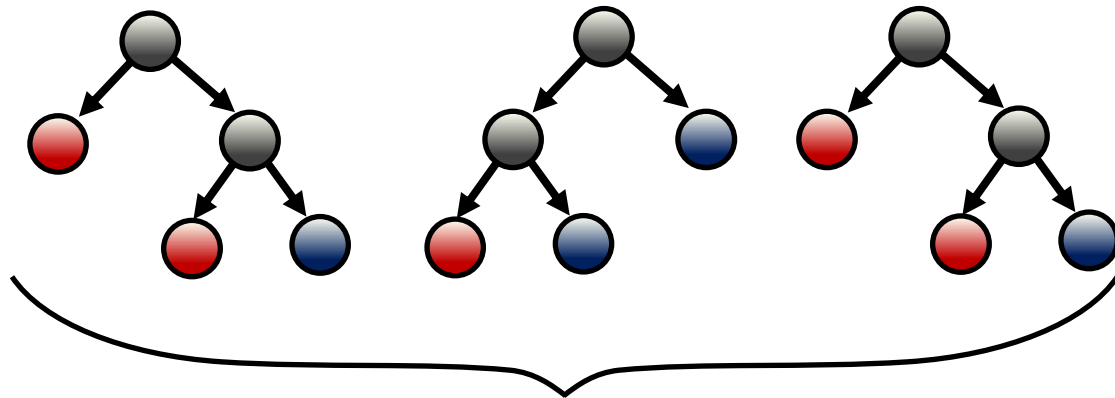
Distribution of Data in Bootstrapped Samples



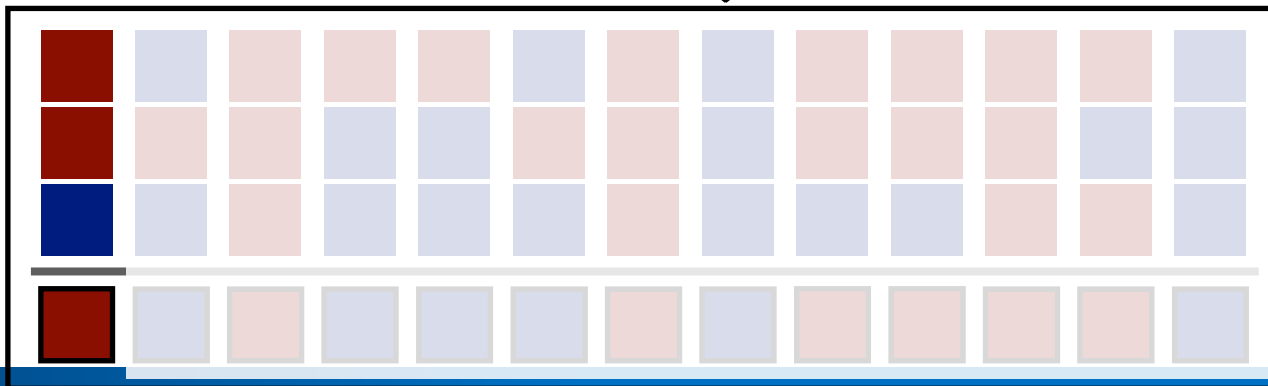
- Given a dataset, create n bootstrapped samples
- For a given record x ,
 $P(\text{rec } x \text{ not selected}) = (1 - 1/n)^n$
- Each bootstrap sample contains approximately $2/3$ of the records

Aggregate Results

Trees vote on or average result for each data point

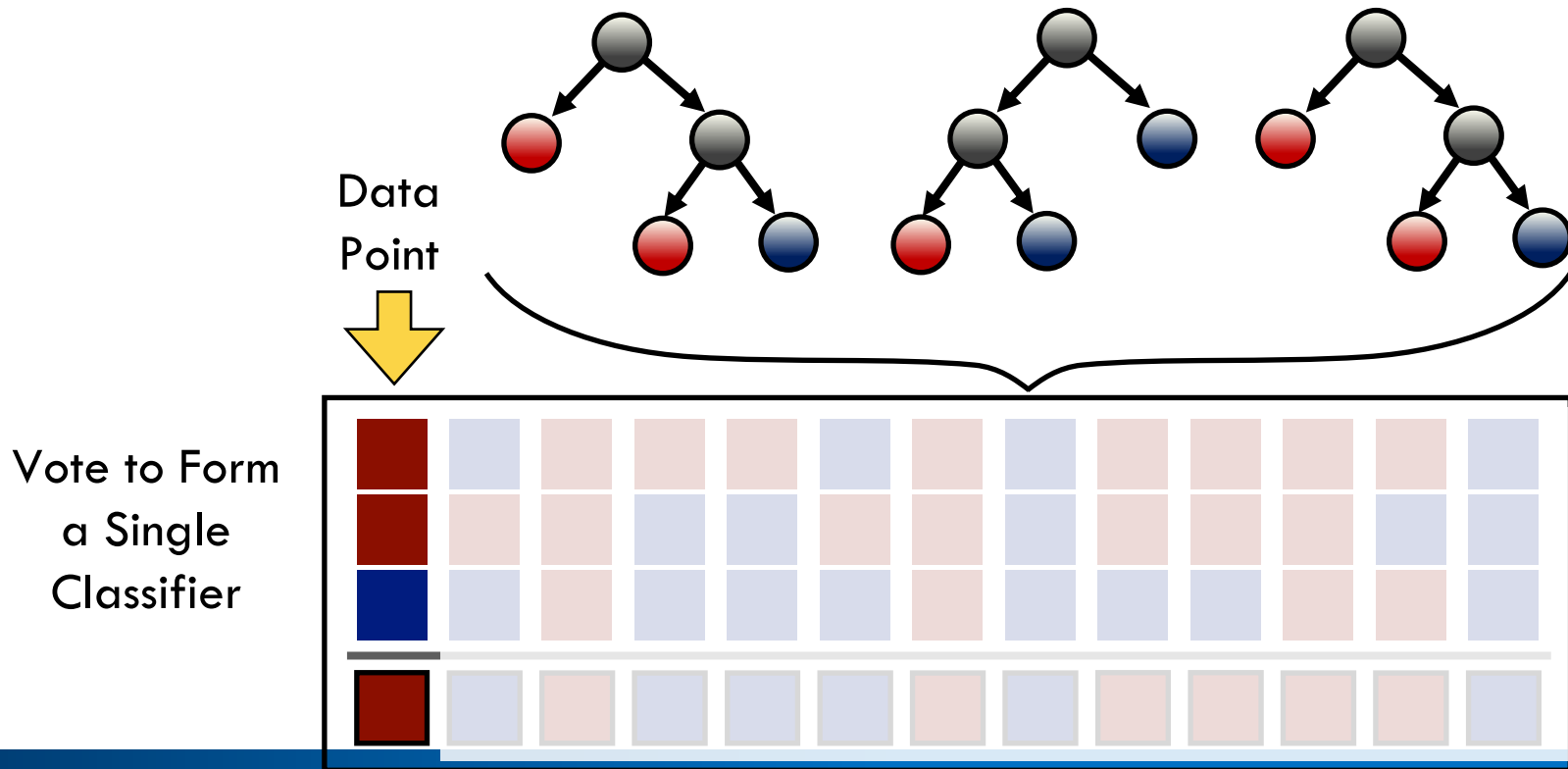


Vote to Form
a Single
Classifier



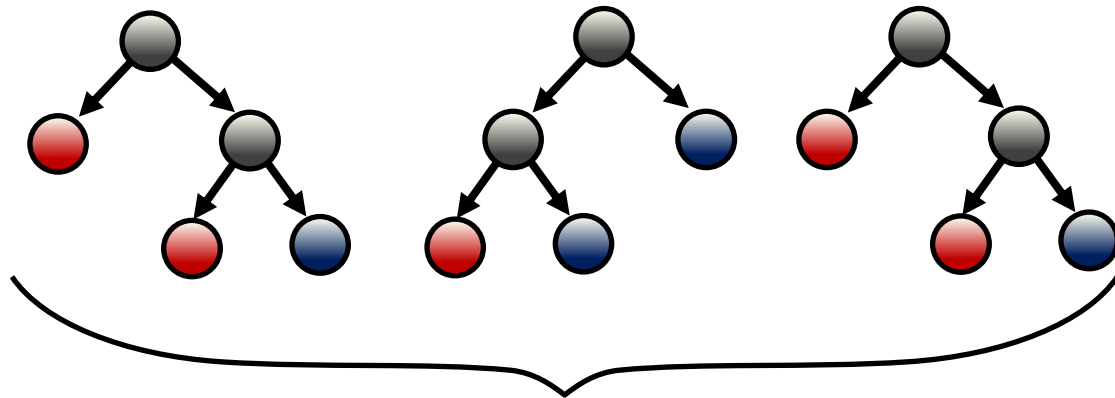
Aggregate Results

Trees vote on or average result for each data point



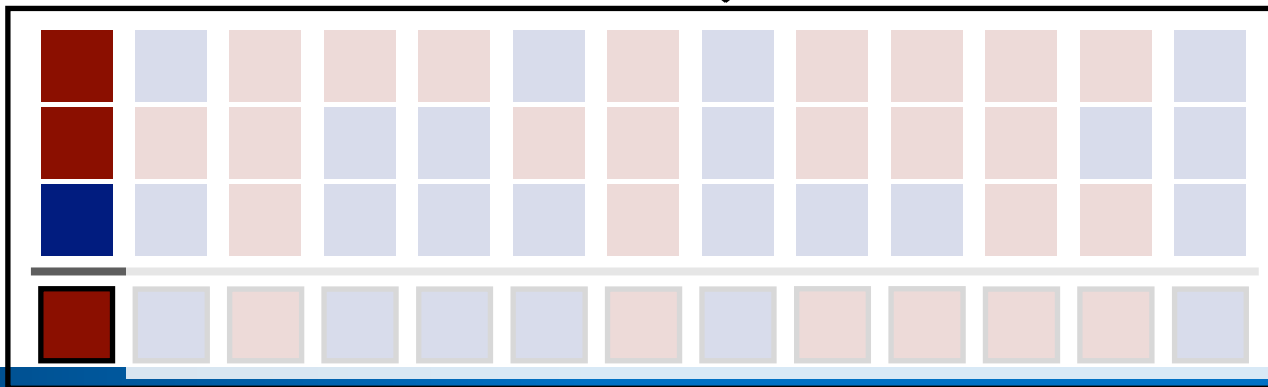
Aggregate Results

Trees vote on or average result for each data point



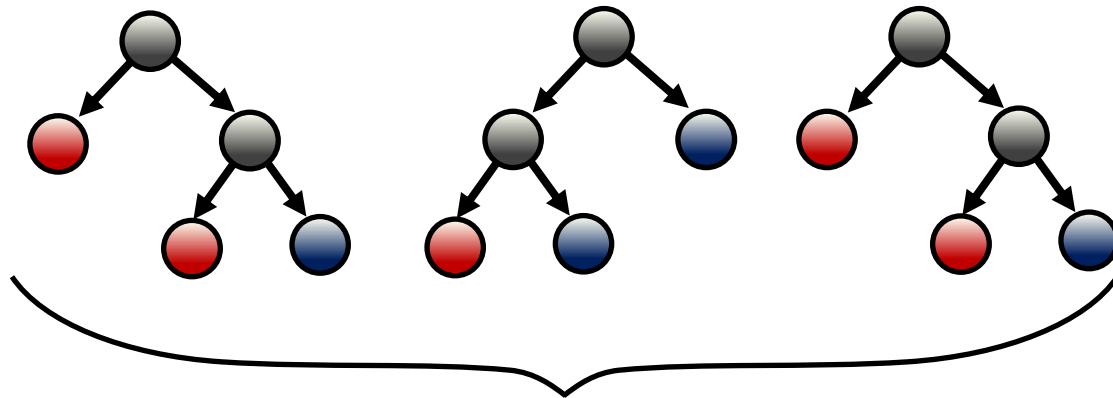
Vote to Form
a Single
Classifier

Results

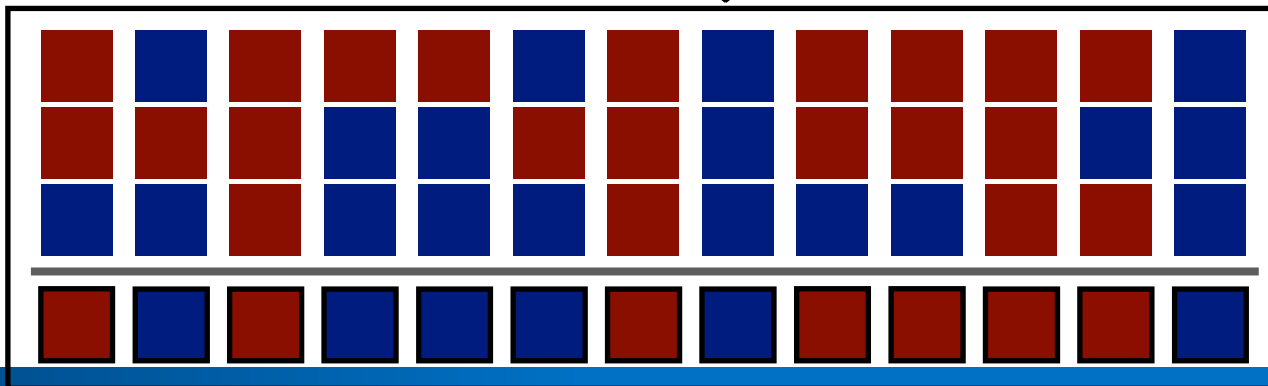


Aggregate Results

Trees vote on or average result for each data point

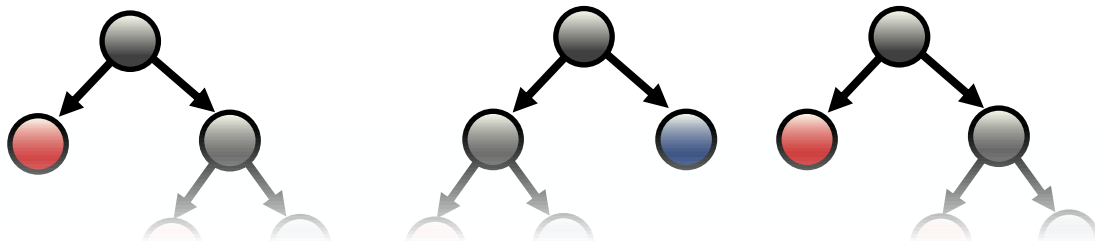


Vote to Form
a Single
Classifier



Aggregate Results

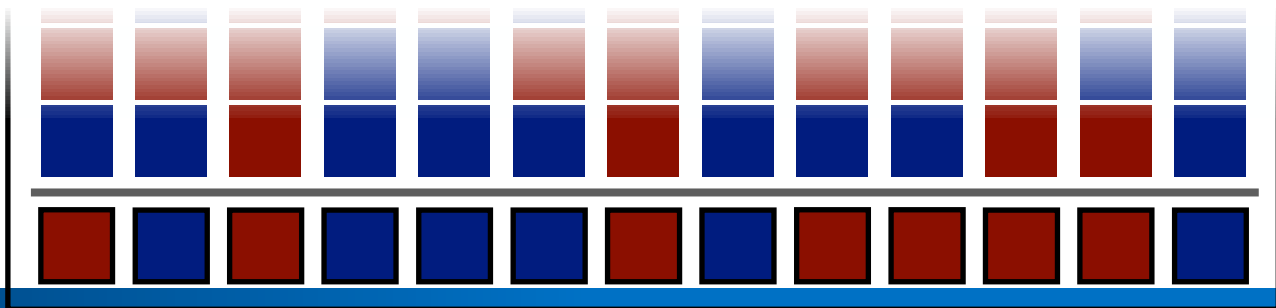
Trees vote on or average result for each data point



Bagging = Bootstrap Aggregating

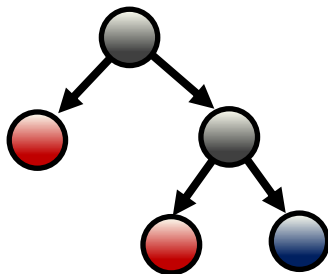
How to Form

a Single
Classifier



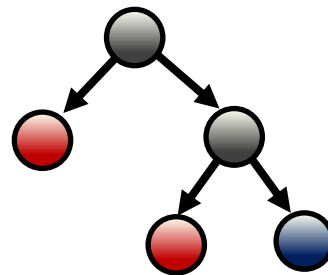
Bagging error calculations

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-29 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



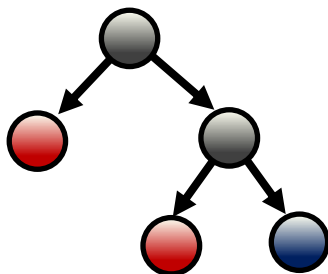
- Bootstrapped samples provide built-in error estimate for each tree
- Create tree based on subset of

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-29 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



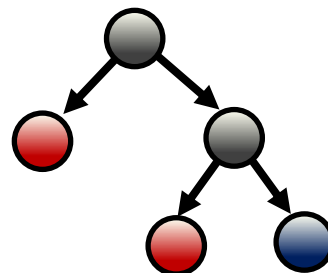
Bagging error calculations

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22	The Hunger Games: Catching Fire	130000000	Francis Lawrence	PG-13	146
1	2013-05-03	Iron Man 3	200000000	Shane Black	PG-13	129
2	2013-11-22	Frozen	150000000	Chris BuckJennifer Lee	PG	108
3	2013-07-03	Despicable Me 2	76000000	Pierre CoffinChris Renaud	PG	98
4	2013-06-14	Man of Steel	225000000	Zack Snyder	PG-13	143
5	2013-10-04	Gravity	100000000	Alfonso Cuaron	PG-13	91
6	2013-06-21	Monsters University	NaN	Dan Scanlon	G	107
7	2013-12-13	The Hobbit: The Desolation of Smaug	NaN	Peter Jackson	PG-13	161
8	2013-05-24	Fast & Furious 6	160000000	Justin Lin	PG-13	130
9	2013-03-08	Oz The Great and Powerful	215000000	Sam Raimi	PG	127
10	2013-05-16	Star Trek Into Darkness	190000000	J.J. Abrams	PG-13	123
11	2013-11-08	Thor: The Dark World	170000000	Alan Taylor	PG-13	120
12	2013-06-21	World War Z	190000000	Marc Forster	PG-13	116
13	2013-03-29	The Croods	135000000	Kirk De MicoChris Sanders	PG	98
14	2013-06-28	The Heat	43000000	Paul Feig	R	117
15	2013-08-07	We're the Millers	37000000	Rawson Marshall Thurber	R	110
16	2013-12-13	American Hustle	40000000	David O. Russell	R	138
17	2013-05-10	The Great Gatsby	105000000	Baz Luhrmann	PG-13	143



- Bootstrapped samples provide built-in error estimate for each tree
- Create tree based on subset of data

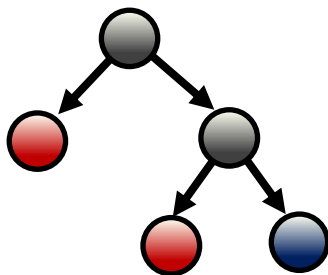
Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22	The Hunger Games: Catching Fire	130000000	Francis Lawrence	PG-13	146
1	2013-05-03	Iron Man 3	200000000	Shane Black	PG-13	129
2	2013-11-22	Frozen	150000000	Chris BuckJennifer Lee	PG	108
3	2013-07-03	Despicable Me 2	76000000	Pierre CoffinChris Renaud	PG	98
4	2013-06-14	Man of Steel	225000000	Zack Snyder	PG-13	143
5	2013-10-04	Gravity	100000000	Alfonso Cuaron	PG-13	91
6	2013-06-21	Monsters University	NaN	Dan Scanlon	G	107
7	2013-12-13	The Hobbit: The Desolation of Smaug	NaN	Peter Jackson	PG-13	161
8	2013-05-24	Fast & Furious 6	160000000	Justin Lin	PG-13	130
9	2013-03-08	Oz The Great and Powerful	215000000	Sam Raimi	PG	127
10	2013-05-16	Star Trek Into Darkness	190000000	J.J. Abrams	PG-13	123
11	2013-11-08	Thor: The Dark World	170000000	Alan Taylor	PG-13	120
12	2013-06-21	World War Z	190000000	Marc Forster	PG-13	116
13	2013-03-29	The Croods	135000000	Kirk De MicoChris Sanders	PG	98
14	2013-06-28	The Heat	43000000	Paul Feig	R	117
15	2013-08-07	We're the Millers	37000000	Rawson Marshall Thurber	R	110
16	2013-12-13	American Hustle	40000000	David O. Russell	R	138
17	2013-05-10	The Great Gatsby	105000000	Baz Luhrmann	PG-13	143



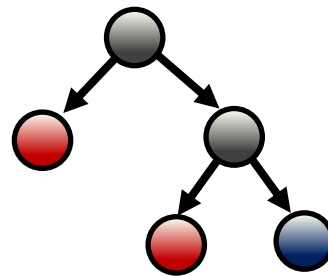
- Measure error for that tree on

Bagging error calculations

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



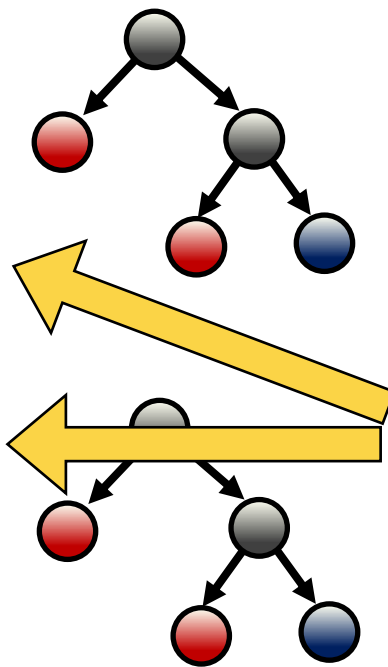
Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



- Bootstrapped samples provide built-in error estimate for each tree
- Create tree based on subset of data
- Measure error for that tree on unused samples
- Called "Out-of-Bag" error

Bagging error calculations

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22	The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13 146
1	2013-05-03	Iron Man 3	200000000	409013994	Shane Black	PG-13 129
2	2013-11-22	Frozen	150000000	400738009	Chris BuckJennifer Lee	PG 108
3	2013-07-03	Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG 98
4	2013-06-14	Man of Steel	225000000	291045518	Zack Snyder	PG-13 143
5	2013-10-04	Gravity	100000000	274092705	Alfonso Cuaron	PG-13 91
6	2013-06-21	Monsters University	NaN	268492764	Dan Scanlon	G 107
7	2013-12-13	The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13 161
8	2013-05-24	Fast & Furious 6	160000000	238679850	Justin Lin	PG-13 130
9	2013-03-08	Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG 127
10	2013-05-16	Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13 123
11	2013-11-08	Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13 120
12	2013-06-21	World War Z	190000000	202359711	Marc Forster	PG-13 116
13	2013-03-29	The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG 98
14	2013-06-28	The Heat	43000000	159582188	Paul Feig	R 117
15	2013-08-07	We're the Millers	37000000	150394119	Rawson Marshall Thurber	R 110
16	2013-12-13	American Hustle	40000000	150117807	David O. Russell	R 138
17	2013-05-10	The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13 143

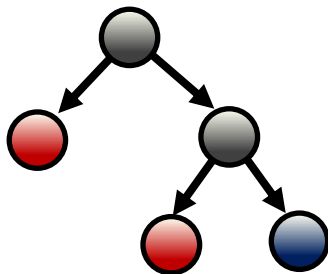


- Bootstrapped samples provide built-in error estimate for each tree
- Create tree based on subset of data
- Measure error for that tree on unused samples

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22	The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13 146
1	2013-05-03	Iron Man 3	200000000	409013994	Shane Black	PG-13 129
2	2013-11-22	Frozen	150000000	400738009	Chris BuckJennifer Lee	PG 108
3	2013-07-03	Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG 98
4	2013-06-14	Man of Steel	225000000	291045518	Zack Snyder	PG-13 143
5	2013-10-04	Gravity	100000000	274092705	Alfonso Cuaron	PG-13 91
6	2013-06-21	Monsters University	NaN	268492764	Dan Scanlon	G 107
7	2013-12-13	The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13 161
8	2013-05-24	Fast & Furious 6	160000000	238679850	Justin Lin	PG-13 130
9	2013-03-08	Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG 127
10	2013-05-16	Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13 123
11	2013-11-08	Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13 120
12	2013-06-21	World War Z	190000000	202359711	Marc Forster	PG-13 116
13	2013-03-29	The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG 98
14	2013-06-28	The Heat	43000000	159582188	Paul Feig	R 117
15	2013-08-07	We're the Millers	37000000	150394119	Rawson Marshall Thurber	R 110
16	2013-12-13	American Hustle	40000000	150117807	David O. Russell	R 138
17	2013-05-10	The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13 143

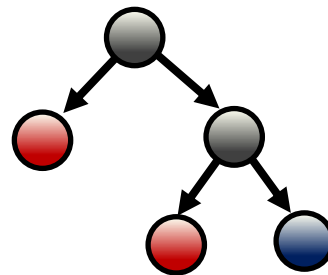
Bagging error calculations

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



- Bootstrapped samples provide built-in error estimate for each tree

- Create tree based on subset of data



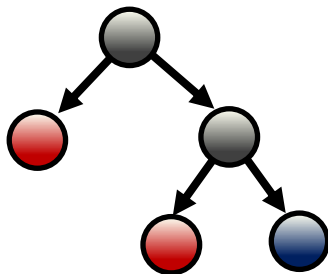
- Measure error for that tree on unused samples

- Called "Out-of-Bag" error

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

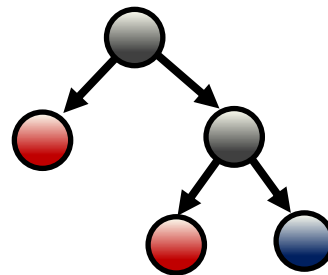
Calculation of Feature Importance

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaH	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



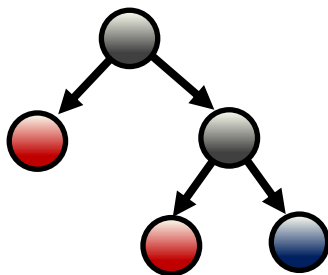
- Fitting a bagged model doesn't produce coefficients like logistic regression

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaH	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



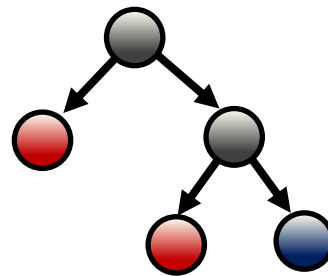
Calculation of Feature Importance

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaH	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



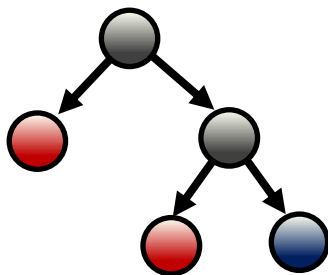
- Fitting a bagged model doesn't produce coefficients like logistic regression
- Instead, feature importances are estimated using oob error

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaH	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



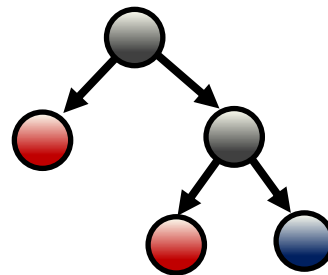
Calculation of Feature Importance

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaH	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



- Fitting a bagged model doesn't produce coefficients like logistic regression
- Instead, feature importances are estimated using oob error

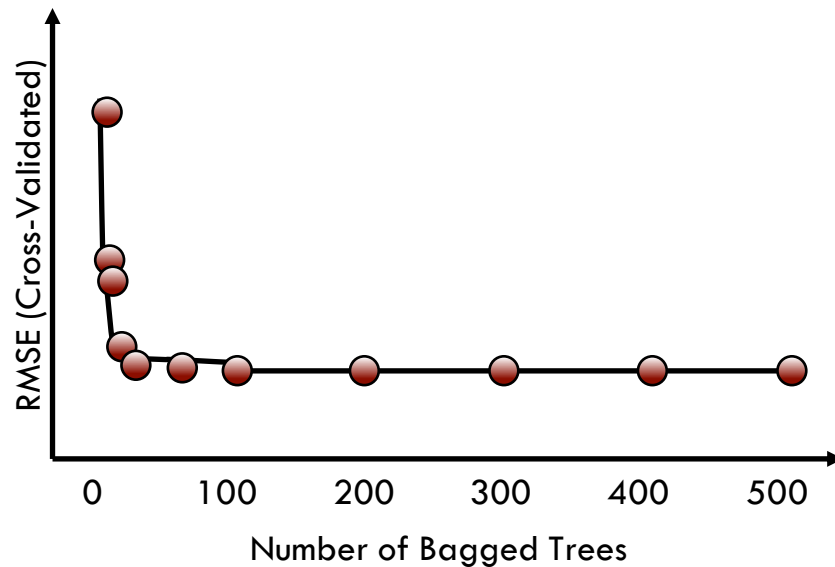
Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaH	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaH	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



- Randomly permute data for particular feature and measure change in accuracy

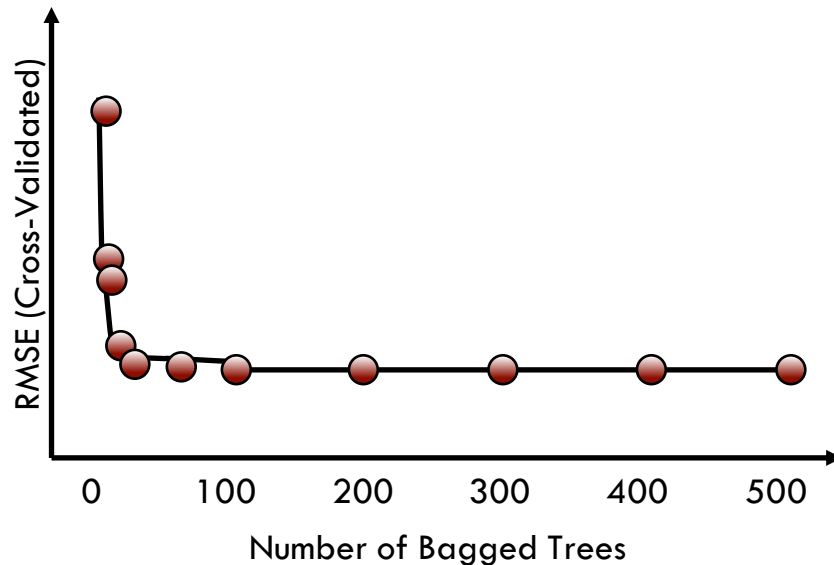
How Many Trees to Fit?

- Bagging performance improvements increase with more trees



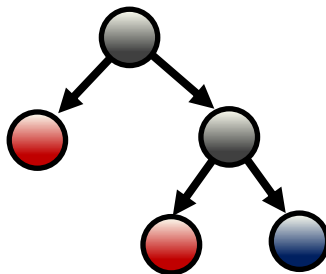
How Many Trees to Fit?

- Bagging performance improvements increase with more trees
- Maximum improvement generally reached ~ 50 trees

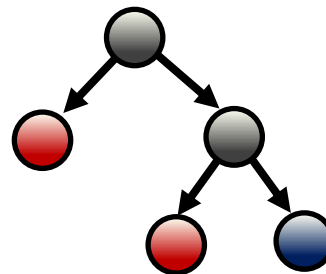


Strengths of Bagging

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-29 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-29 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

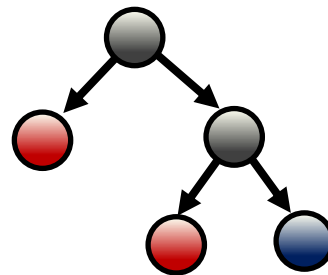
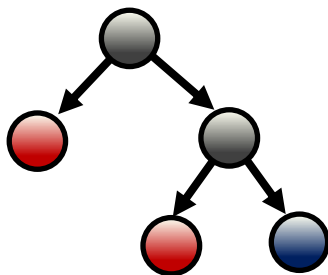


Same as decision trees:

- Easy to interpret and implement

Strengths of Bagging

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



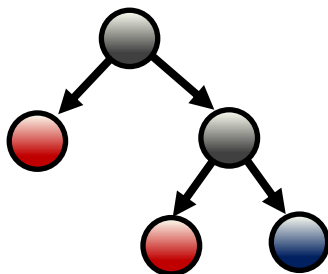
Same as decision trees:

- Easy to interpret and implement
- Heterogeneous input data allowed, no preprocessing required

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

Strengths of Bagging

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	42468047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

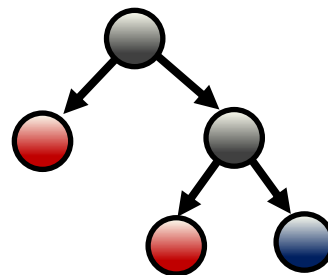


Same as decision trees:

- Easy to interpret and implement
- Heterogeneous input data allowed, no preprocessing required

Specific to bagging:

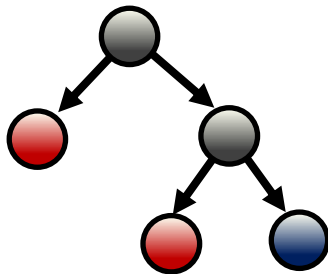
- Less variability than decision trees



Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	42468047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

Strengths of Bagging

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-29 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



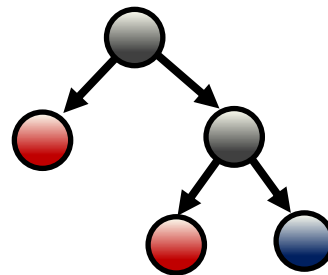
Same as decision trees:

- Easy to interpret and implement
- Heterogeneous input data allowed, no preprocessing required

Specific to bagging:

- Less variability than decision trees
- Can grow trees in parallel

Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22 The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03 Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22 Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03 Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14 Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04 Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21 Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13 The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24 Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08 Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16 Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08 Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21 World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-29 The Croods	135000000	187168425	Kirk De MicoChris Sanders	PG	98
14	2013-06-28 The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07 We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13 American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10 The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143



BaggingClassifier: The Syntax

Import the class containing the classification method

```
from sklearn.ensemble import BaggingClassifier
```

BaggingClassifier: The Syntax

Import the class containing the classification method

```
from sklearn.ensemble import BaggingClassifier
```

Create an instance of the class

```
BC = BaggingClassifier(n_estimators=50)
```

BaggingClassifier: The Syntax

Import the class containing the classification method

```
from sklearn.ensemble import BaggingClassifier
```

Create an instance of the class

```
BC = BaggingClassifier(n_estimators=50)
```

Fit the instance on the data and then predict the expected value

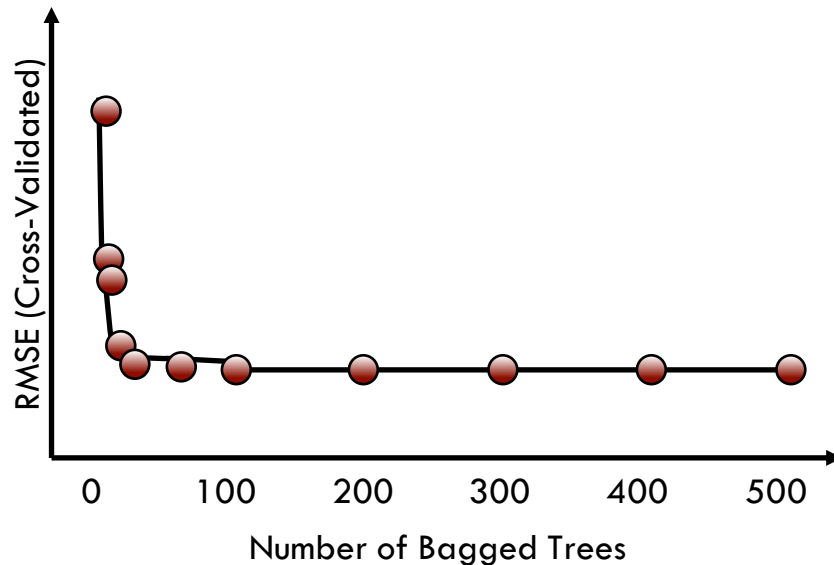
```
BC = BC.fit(X_train, y_train)  
y_predict = BC.predict(X_test)
```

Tune parameters with cross-validation. Use BaggingRegressor for regression.

Reduction in Variance Due to Bagging

- For n independent trees, each with variance σ^2 , the bagged variance is:

$$\frac{\sigma^2}{n}$$



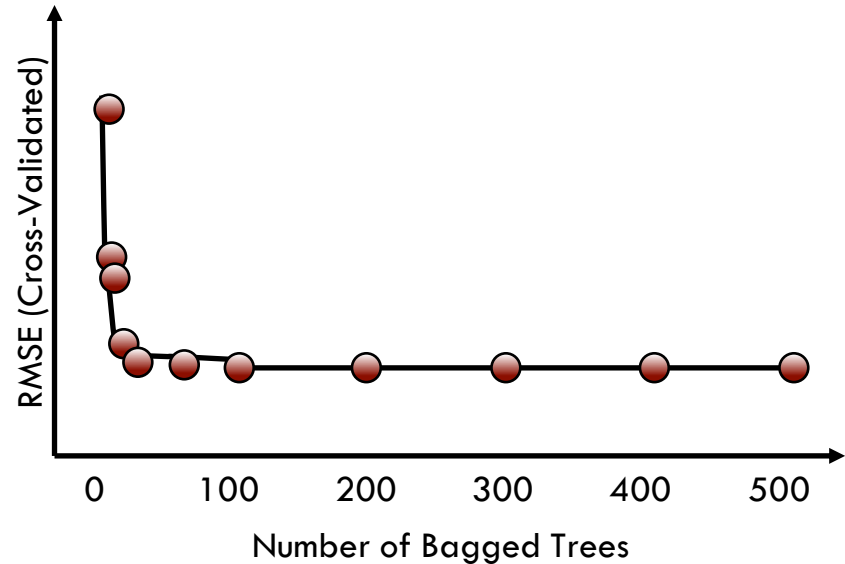
Reduction in Variance Due to Bagging

- For n independent trees, each with variance σ^2 , the bagged variance is:

$$\frac{\sigma^2}{n}$$

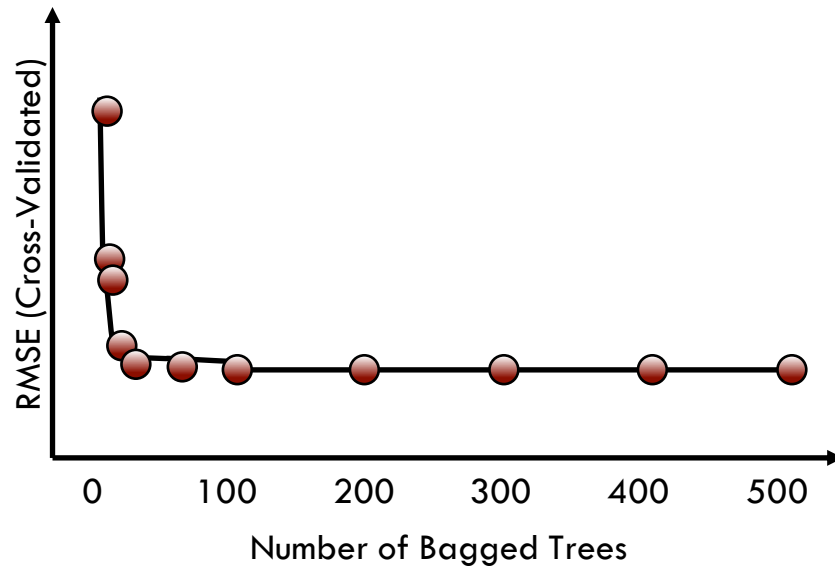
- However, bootstrap samples are correlated (ρ):

$$\rho\sigma^2 + \frac{1-\rho}{n}\sigma^2$$



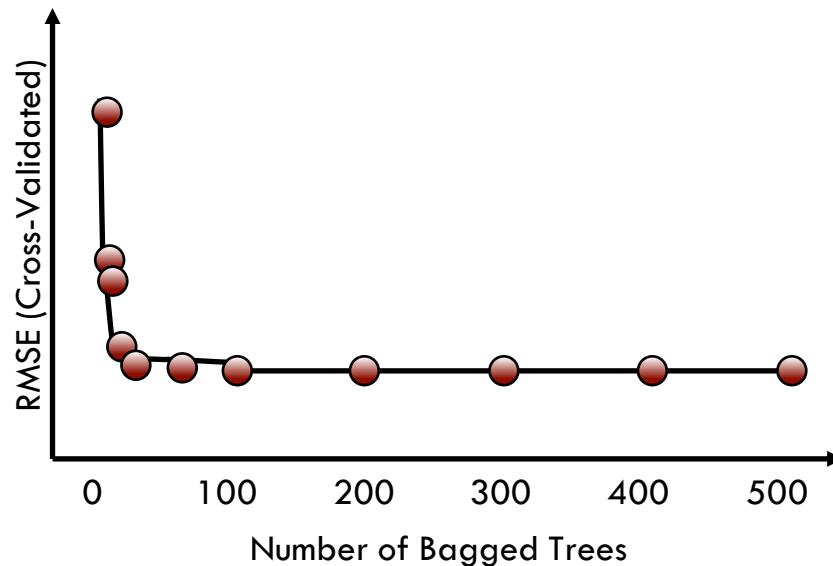
Introducing More Randomness

- Solution: further de-correlate trees



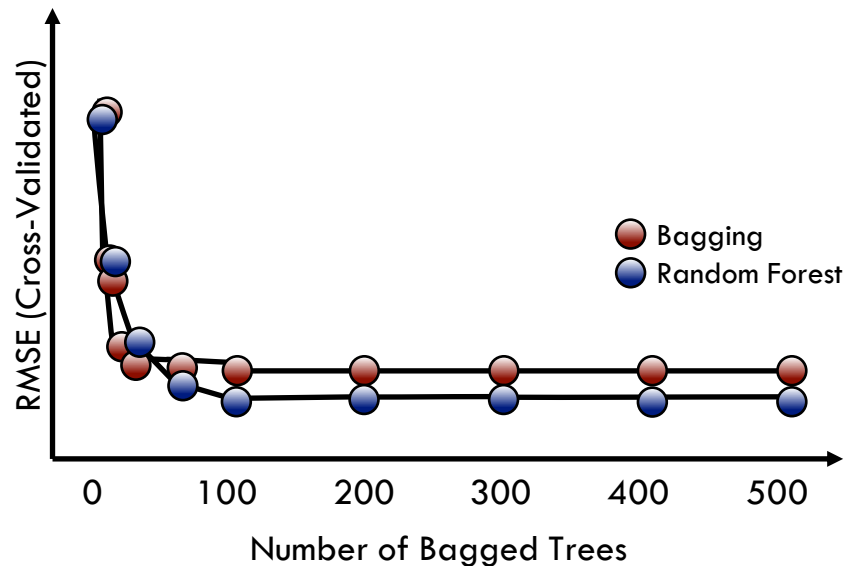
Introducing More Randomness

- Solution: further de-correlate trees
- Use random subset of features for each tree
 - Classification: \sqrt{m}
 - Regression: $m/3$



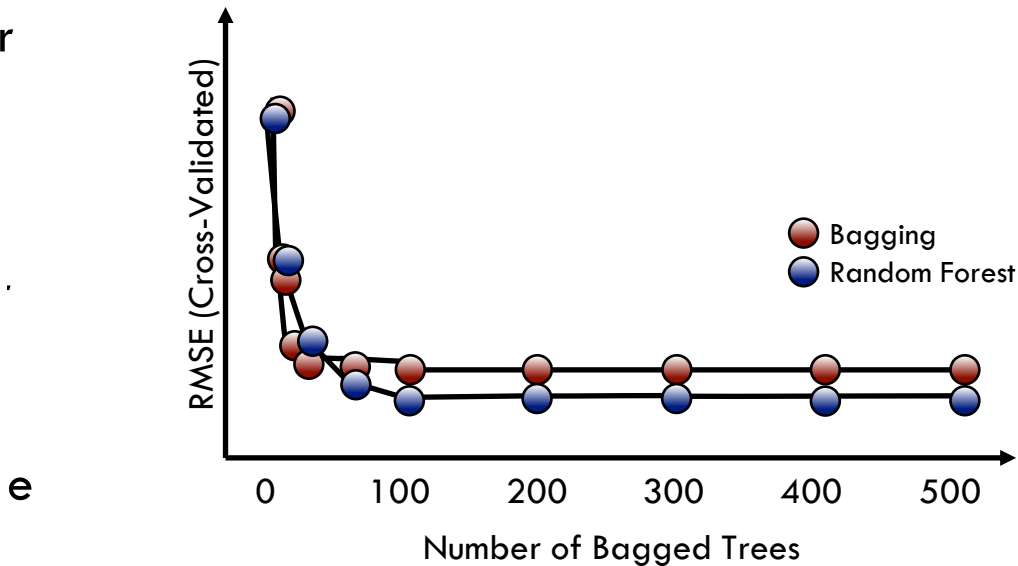
Introducing More Randomness

- Solution: further de-correlate trees
- Use random subset of features for each tree
 - Classification: \sqrt{m}
 - Regression: $m/3$
- Called "Random Forest"



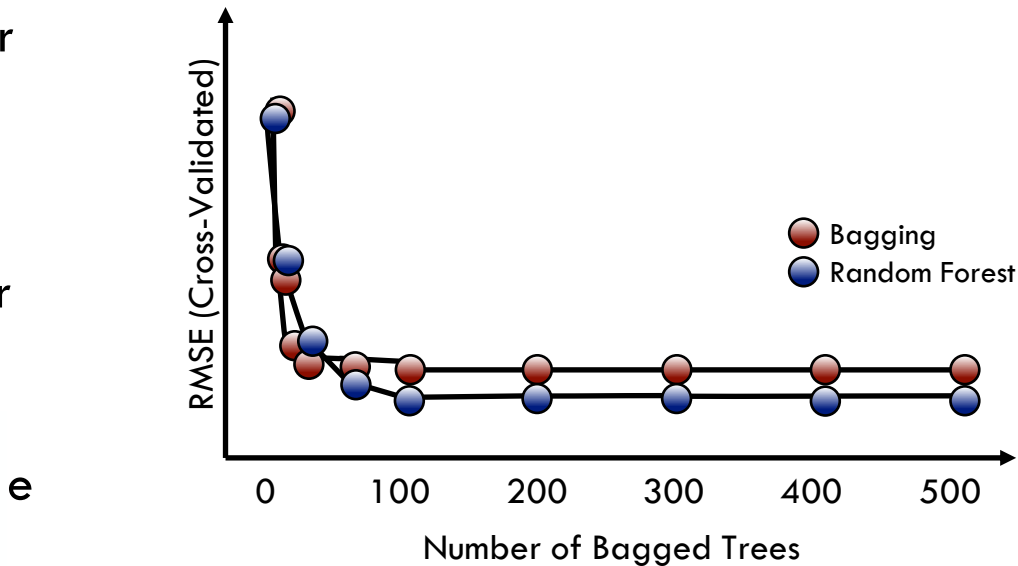
How Many Random Forest Trees?

- Errors are further reduced for Random Forest relative to Bagging



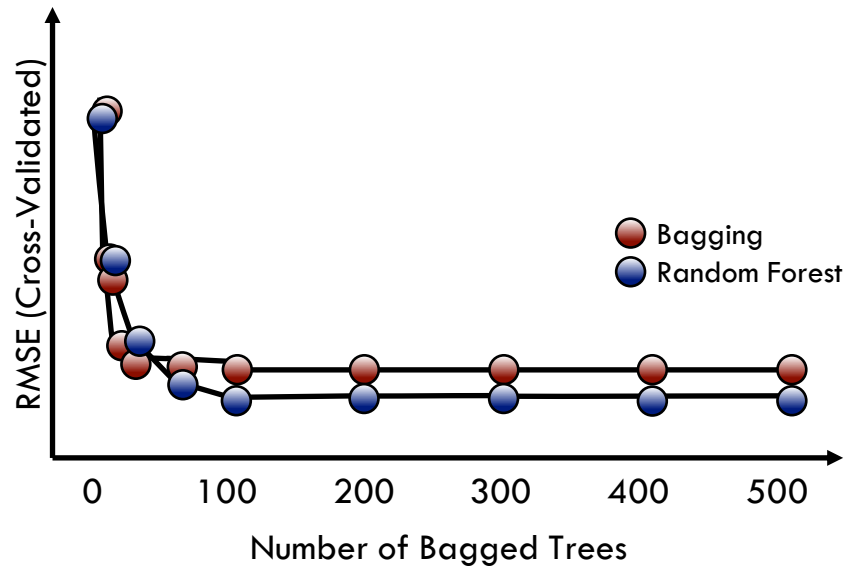
How Many Random Forest Trees?

- Errors are further reduced for Random Forest relative to Bagging
- Grow enough trees until error settles down



How Many Random Forest Trees?

- Errors are further reduced for Random Forest relative to Bagging
- Grow enough trees until error settles down
- Additional trees won't improve results



RandomForest: The Syntax

Import the class containing the classification method

```
from sklearn.ensemble import RandomForestClassifier
```

To use the Intel® Extension for Scikit-learn* variant of this algorithm:

- Install Intel® oneAPI AI Analytics Toolkit (AI Kit)
- **Add the following two lines of code after the above code:**

```
import patch_sklearn  
patch_sklearn()
```

RandomForest: The Syntax

Import the class containing the classification method

```
from sklearn.ensemble import RandomForestClassifier
```

Create an instance of the class

```
RC = RandomForestClassifier(n_estimators=100, max_features=10)
```

Fit the instance on the data and then predict the expected value

```
RC = RC.fit(X_train, y_train)  
y_predict = RC.predict(X_test)
```

Tune parameters with cross-validation. Use `RandomForestRegressor` for regression.

Introducing Even More Randomness

- Sometimes additional randomness is desired beyond Random Forest
- Solution: select features randomly and create splits randomly—

Introducing Even More Randomness

- Sometimes additional randomness is desired beyond Random Forest
- Solution: select features randomly and create splits randomly—
don't choose greedily
- Called "Extra Random Trees"

Introducing Even More Randomness

- Sometimes additional randomness is desired beyond Random Forest
- Solution: select features randomly and create splits randomly—
don't choose greedily
- Called "Extra Random Trees"

ExtraTreesClassifier: The Syntax

Import the class containing the classification method

```
from sklearn.ensemble import ExtraTreesClassifier
```

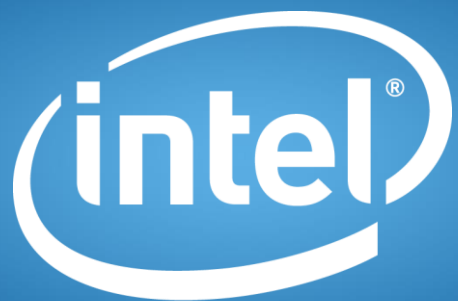
Create an instance of the class

```
EC = ExtraTreesClassifier(n_estimators=100, max_features=10)
```

Fit the instance on the data and then predict the expected value

```
EC = EC.fit(X_train, y_train)  
y_predict = EC.predict(X_test)
```

Tune parameters with cross-validation. Use `ExtraTreesRegressor` for regression.



Software