Random Forests:
Theory & Intuition

Bias-Variance Tradeoff

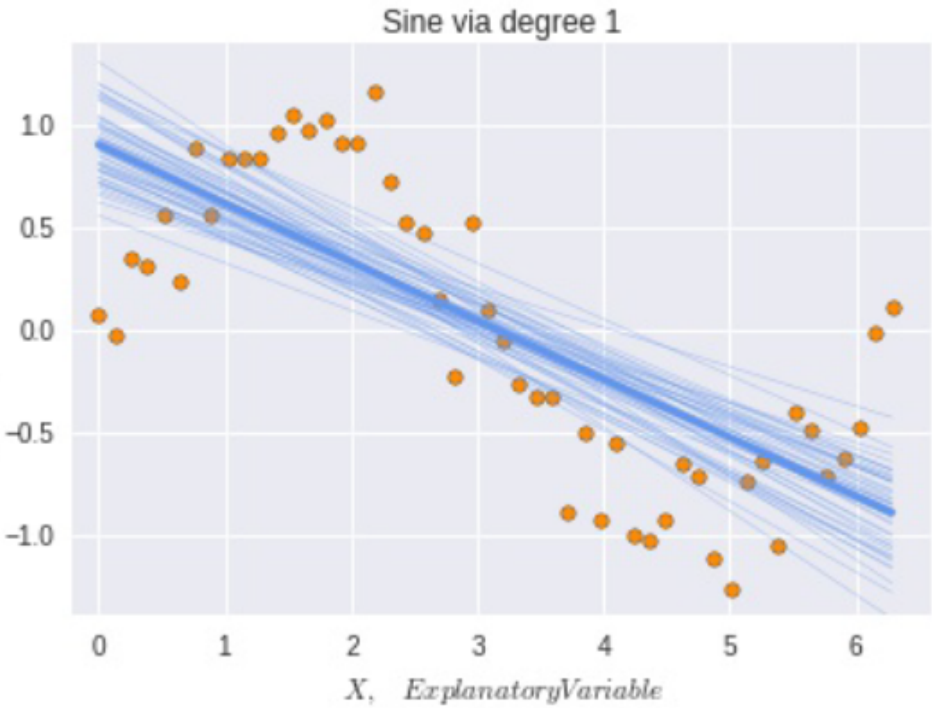
$$E[mean\ squared\ error] = noise + \ bias^2 + variance$$

ias: Difference between averaged prediction of model versus the Bayes mode
ariance: Variability of learned model conditioned upon the learning set.
oise: Irreducible error.

Bias-Variance Tradeoff

$$E[mean\ squared\ error] = noise + \ bias^2 + variance$$

Bias: Difference between averaged prediction of model versus the Bayes mode
Variance: Variability of learned model conditioned upon the learning set.
Noise: Irreducible error.

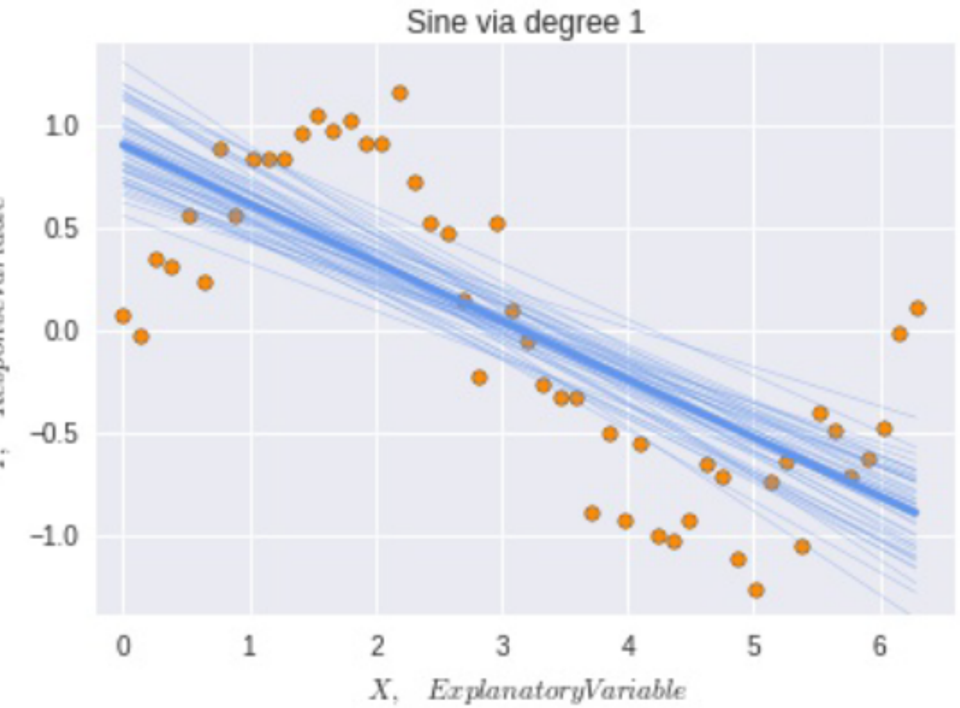


Underfit model: High Bias, Low Variance.

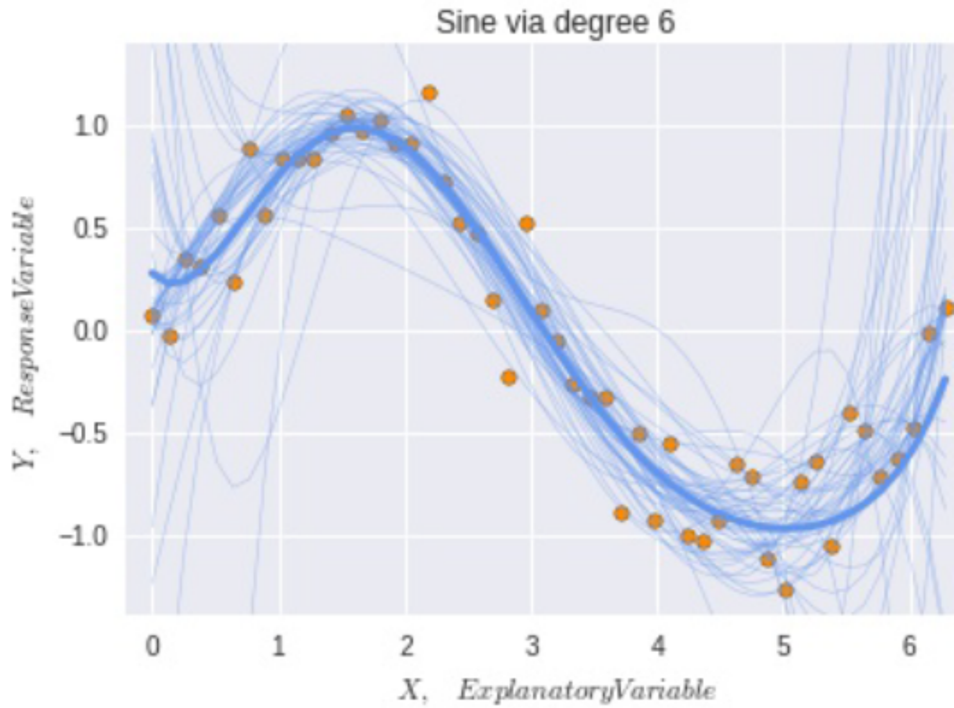
Bias-Variance Tradeoff

$$E[mean\ squared\ error] = noise + \ bias^2 + variance$$

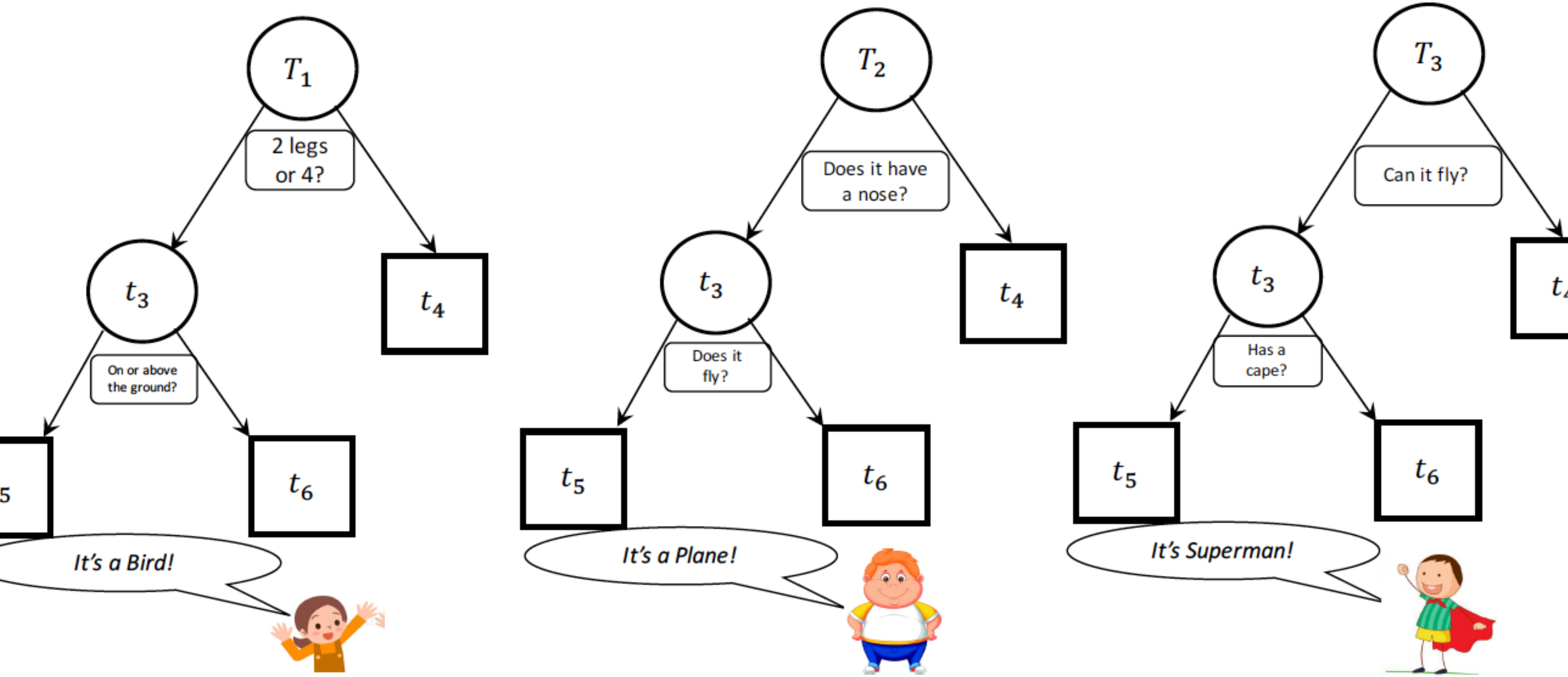
Bias: Difference between averaged prediction of model versus the Bayes mod
Variance: Variability of learned model conditioned upon the learning set.
Noise: Irreducible error.



Underfit model: High Bias, Low Variance.



Overfit model: High Variance, Low Bias.

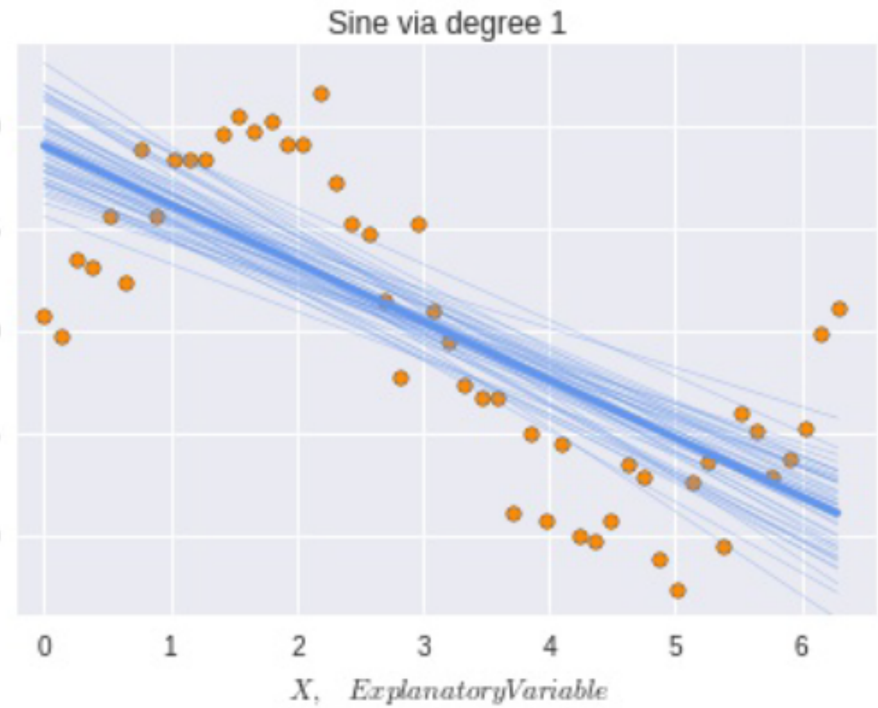


T_1
2 legs or 4?
t_3
t_4
On or above the ground?
t_6
It's a Bird!
T_2
Does it have a nose?
t_3
t_4
Does it fly?
t_5
t_6
It's a Plane!
T_3
Can it fly?
t_3
Has a cape?
t_5
t_6
It's Superman!

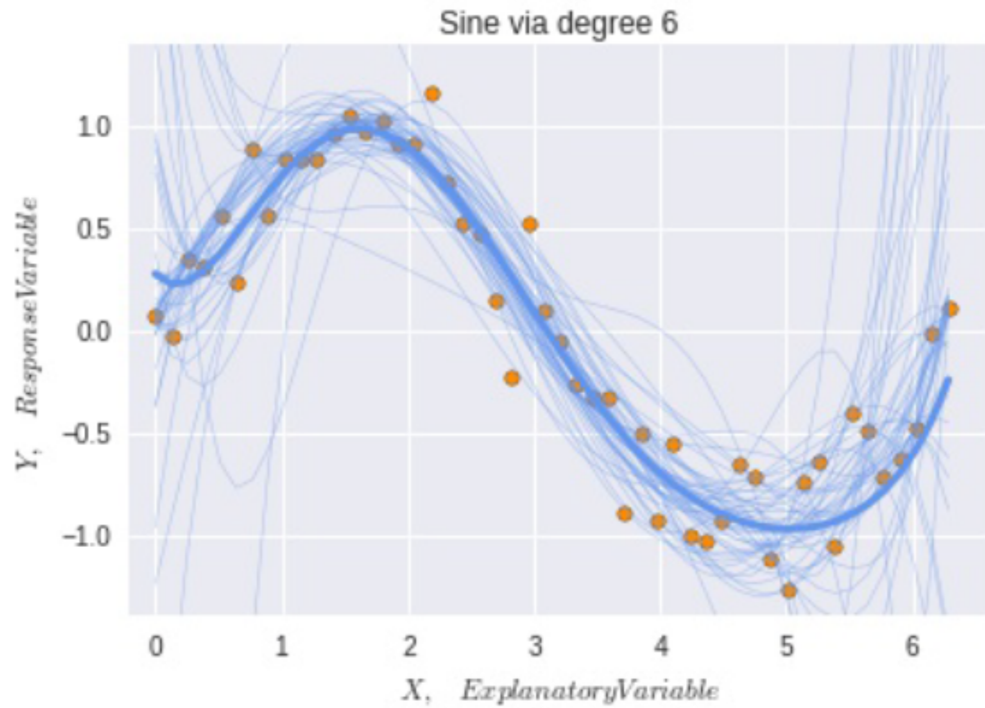
Bias-Variance Tradeoff

$$E[mean\ squared\ error] = noise + \ bias^2 + variance$$

s: Difference between averaged prediction of model versus the Bayes model.
iance: Variability of learned model conditioned upon the learning set.
se: Irreducible error.



Underfit model: High Bias, Low Variance.



Overfit model: High Variance, Low Bias.

eraging over large ensembles of decorrelated models reduces variance, while maintaining bias.

Ensemble Learning

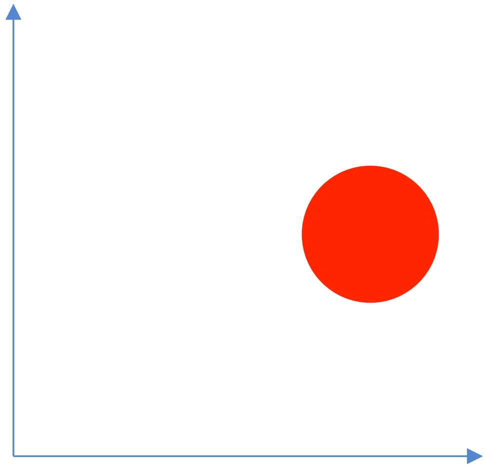
ondorcet's Jury Theorem: A group wants to arrive at the "correct" decision via
ajority vote, wherein each individual has a probability p of voting for the
orrect decision. What should the size of the group be for optimal performanc

ntuitive example: The performance of the "Ask The Audience" lifeline in Who
Vants To Be A Millionaire? (*92%* accuracy vs *65%* for "Phone A Friend")

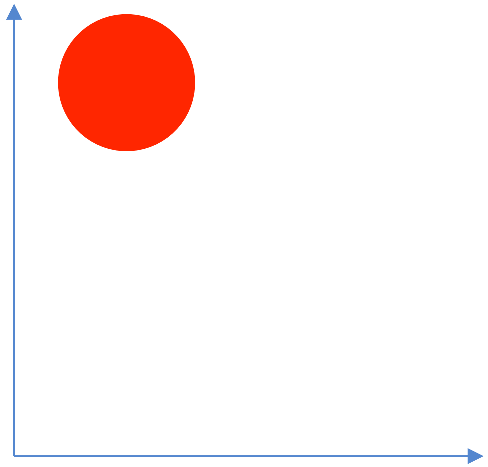
ggregating randomized models decreases the variance of the ensemble.

nsemble Prediction: Weather forecasting.

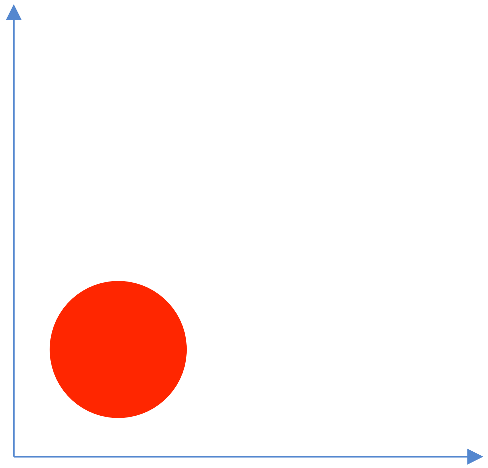
Ensemble Learning: Intuition



Model 1

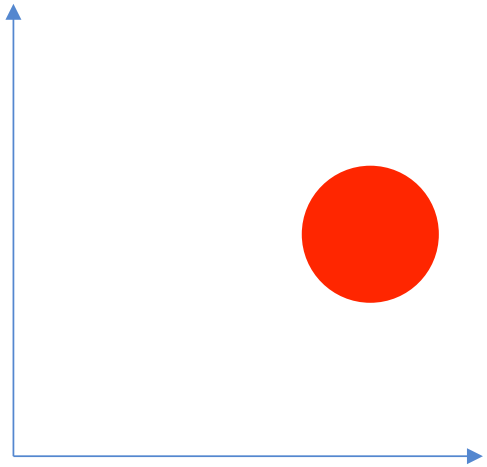


Model 2

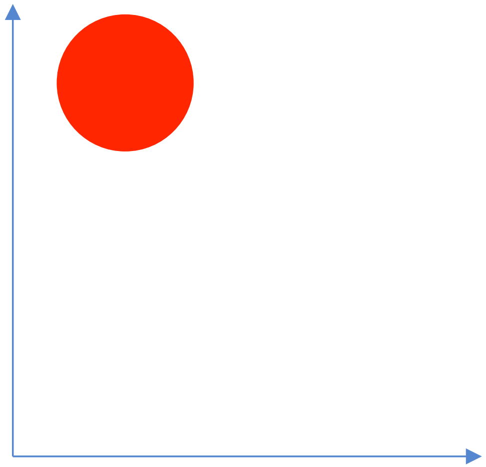


Model 3

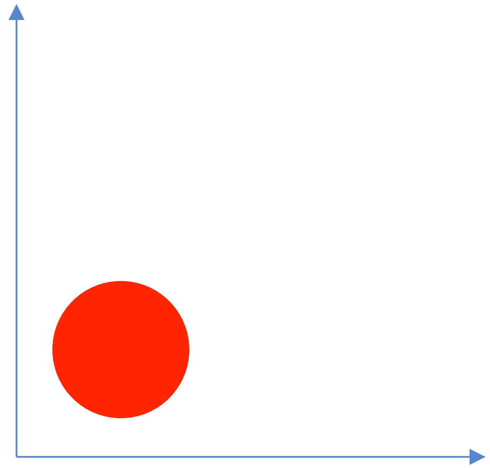
Ensemble Learning: Intuition



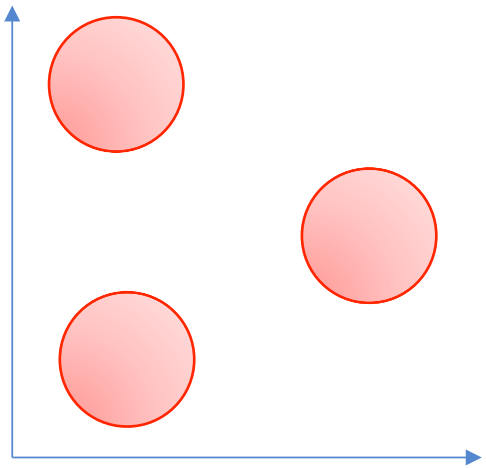
Model 1



Model 2



Model 3



Model Ensemble

Random Forests: Details

- Bagging: **B**ootstrap **agg**regation

- Random Subsampling over features (aka “feature bagging”)

Random Forests: Bootstrapping

Original Toy Dataset

1	X2	X3	X4	Y
	0	2.7	314	17
	0	3.14	516	42
	1	3.14	400	31
	1	2.3	300	99
	0	2.3	200	157

Random Forests: Bootstrapping

Original Toy Dataset

X2	X3	X4	Y
0	2.7	314	17
0	3.14	516	42
1	3.14	400	31
1	2.3	300	99
0	2.3	200	157

Bootstrapped Dataset 1

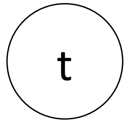
X1	X2	X3	X4	Y
1	0	2.7	314	17
1	0	2.7	314	17
3	1	3.14	400	31
5	1	2.3	300	99
5	1	2.3	300	99

Random Forests: Feature Bagging

Bootstrapped Dataset 1

1	X2	X3	X4	Y
	0	2.7	314	17
	0	2.7	314	17
	1	3.14	400	31
	1	2.3	300	99
	1	2.3	300	99

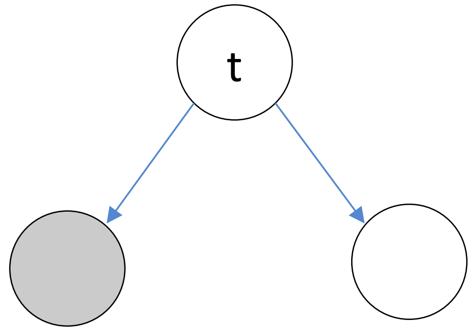
Tree 1



Random Forests: Feature Bagging

1	X2	X3	X4	Y
	0	[redacted]	[redacted]	17
	0	[redacted]	[redacted]	17
	1	[redacted]	[redacted]	31
	1	[redacted]	[redacted]	99
	1	[redacted]	[redacted]	99

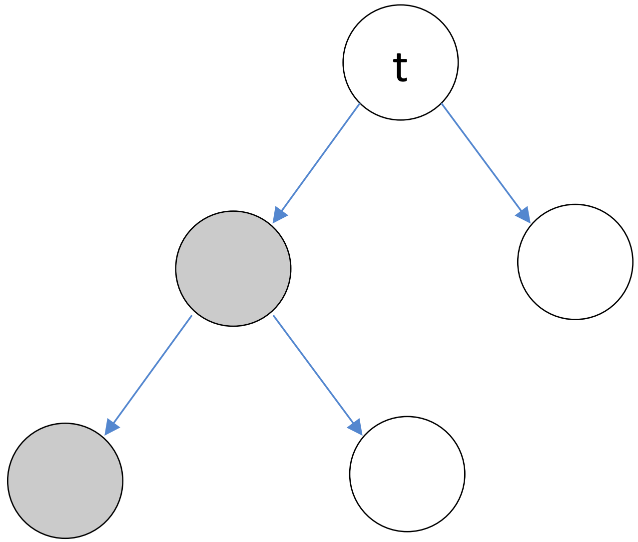
Tree 1



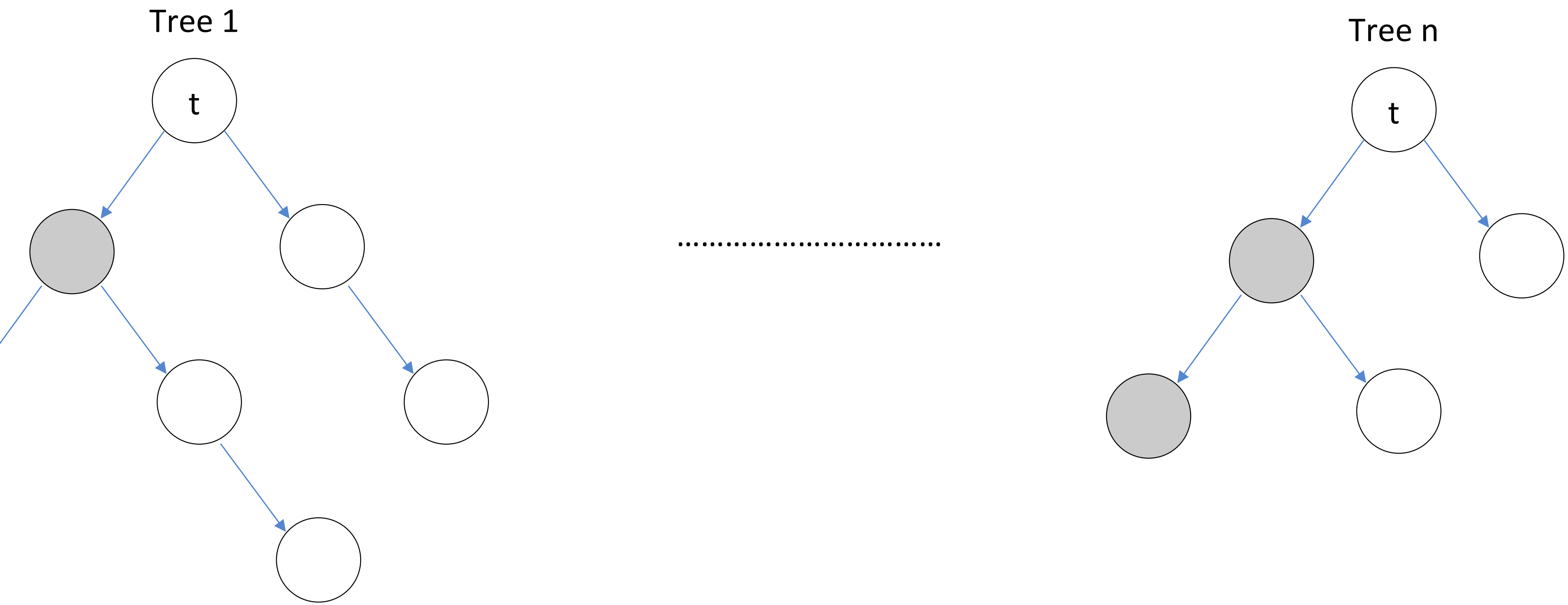
Random Forests: Feature Bagging

1	X2	X3	X4	Y
	0	2.7		17
	0	2.7		17
	1	3.14		31
	1	2.3		99
	1	2.3		99

Tree 1



Random Forests: Ensemble

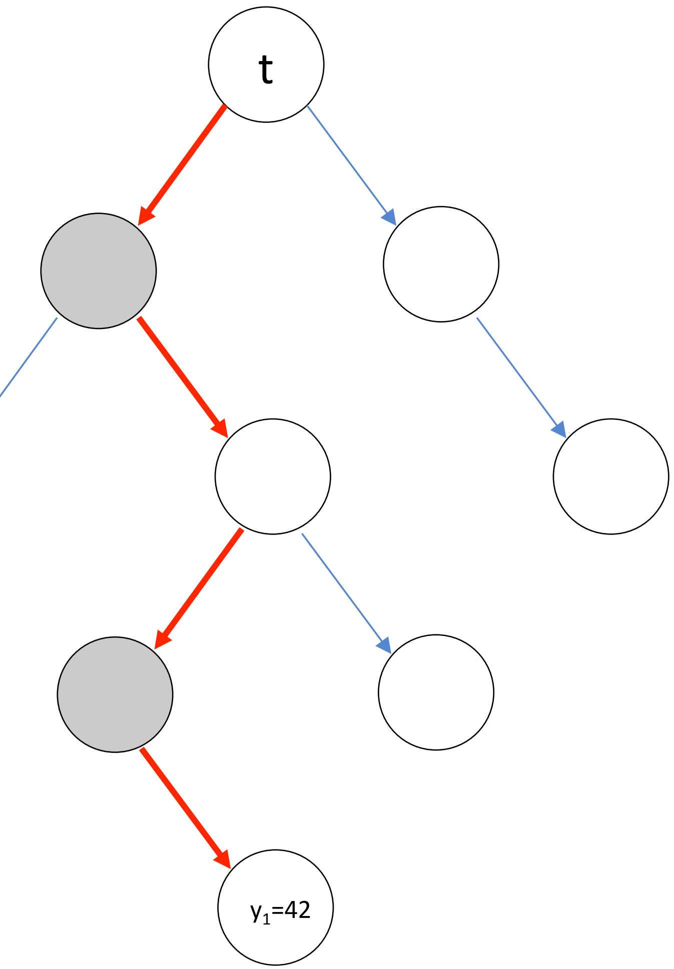


Random Forests: Predictions (Aggregation)

New Sample for Prediction

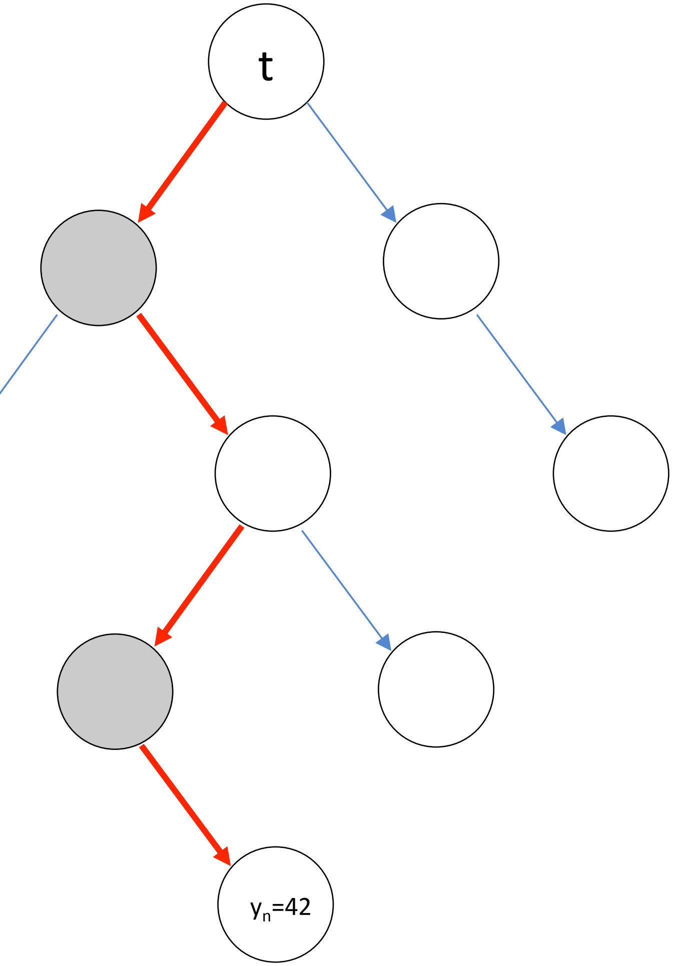
X1	X2	X3	X4
1	0	2.7	314

Tree 1



Random Forests: Predictions (Aggregation)

Tree n



New Sample for Prediction

X1	X2	X3	X4
1	0	2.7	314

Regression, $y_{prediction} = y_1 + y_2 + y$

Classification, $y_{prediction} = \text{majority}(y_i)$

Random Forests: Easy Error Estimation, OOB estimates

Original Toy Dataset

1	X2	X3	X4	Y
	0	2.7	314	17
	0	3.14	516	42
	1	3.14	400	31
	1	2.3	300	99
	0	2.3	200	157

Bootstrapped Dataset 1

X1	X2	X3	X4	Y
1	0	2.7	314	17
1	0	2.7	314	17
3	1	3.14	400	31
5	1	2.3	300	99
5	1	2.3	300	99

n average, about 32% of the samples will not end up in a specific Bootstrapped dataset.
se are referred to as Out Of Bag samples, or OOB samples.

Random Forests: Easy Error Estimation, OOB estimates

ep 1: For every sample, determine the trees where it is an OOB sample.

ep 2: Make predictions for this sample from these trees.

ep 3: Calculate final averaged prediction for this sample.

ep 4: Calculate error for this sample.

ep 5: Repeat for all points to find OOB error.