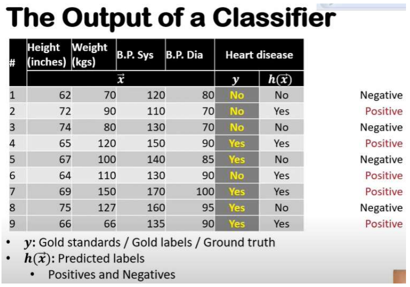
ML Evaluation

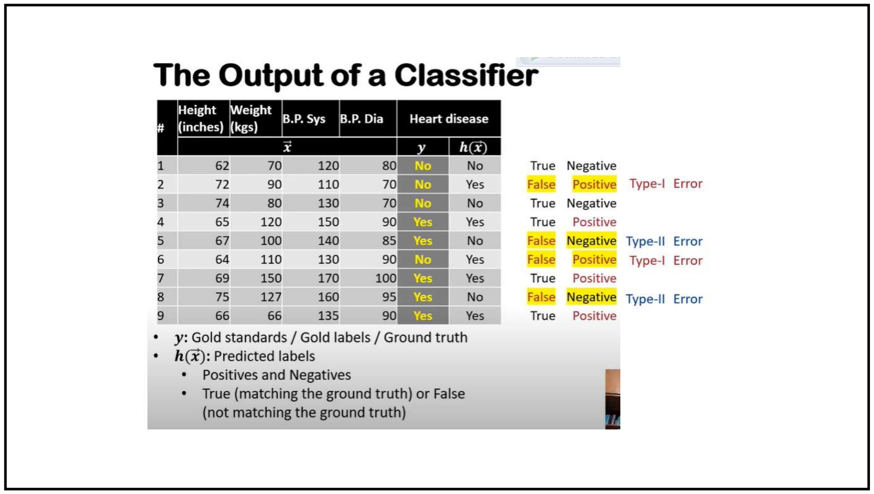
Courtesy Dr. Agha Ali Raza



The Output of a Classifier

#	Height (inches)	Weight (kgs)	B.P. Sys	B.P. Dia	Heart disease		
	$\vec{x}$				y	$h(\vec{x})$	
1	62	70	120	80	No	No	Negative
2	72	90	110	70	No	Yes	Positive
3	74	80	130	70	No	No	Negative
4	65	120	150	90	Yes	Yes	Positive
5	67	100	140	85	Yes	No	Negative
6	64	110	130	90	No	Yes	Positive
7	69	150	170	100	Yes	Yes	Positive
8	75	127	160	95	Yes	No	Negative
9	66	66	135	90	Yes	Yes	Positive

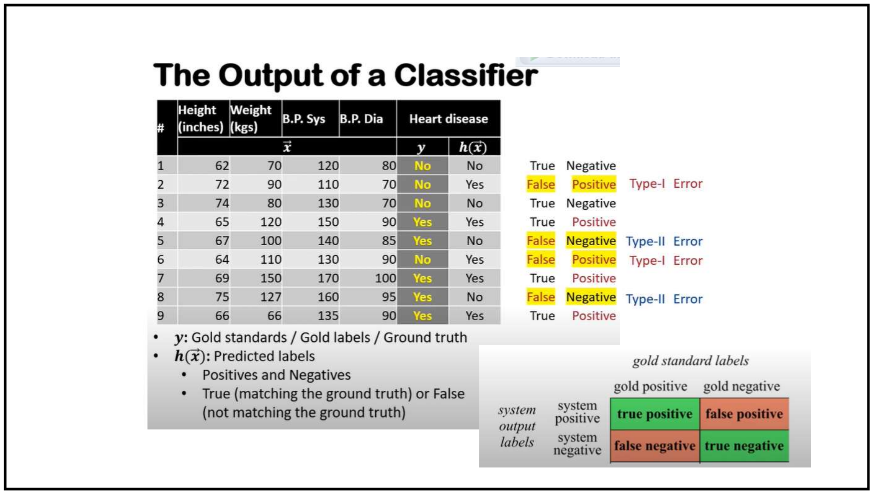
- y: Gold standards / Gold labels / Ground truth
- $h(\vec{x})$: Predicted labels
  - Positives and Negatives



The Output of a Classifier

#	Height (inches)	Weight (kgs)	B.P. Sys	B.P. Dia	Heart disease				
	$\vec{x}$				y	$h(\vec{x})$			
1	62	70	120	80	No	No	True	Negative	
2	72	90	110	70	No	Yes	False	Positive	Type-I Error
3	74	80	130	70	No	No	True	Negative	
4	65	120	150	90	Yes	Yes	True	Positive	
5	67	100	140	85	Yes	No	False	Negative	Type-II Error
6	64	110	130	90	No	Yes	False	Positive	Type-I Error
7	69	150	170	100	Yes	Yes	True	Positive	
8	75	127	160	95	Yes	No	False	Negative	Type-II Error
9	66	66	135	90	Yes	Yes	True	Positive	

- y**:** Gold standards / Gold labels / Ground truth
- $h(\vec{x})$**:** Predicted labels
  - Positives and Negatives
  - True (matching the ground truth) or False (not matching the ground truth)

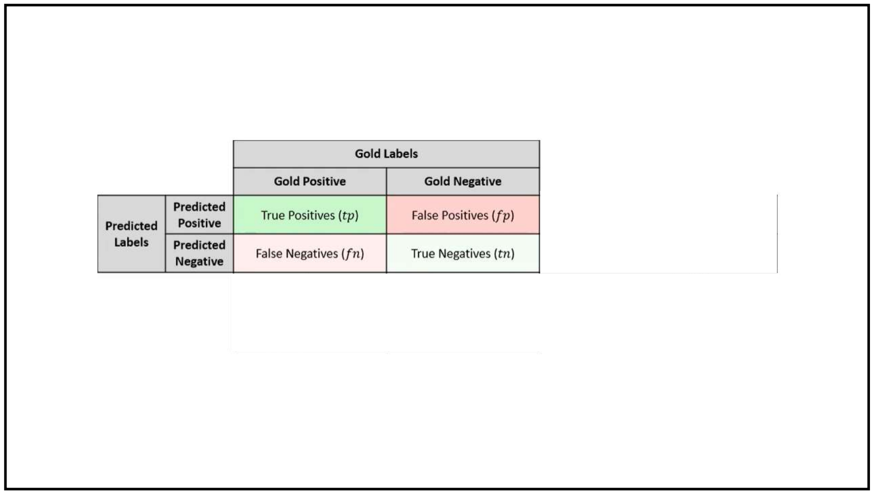


The Output of a Classifier

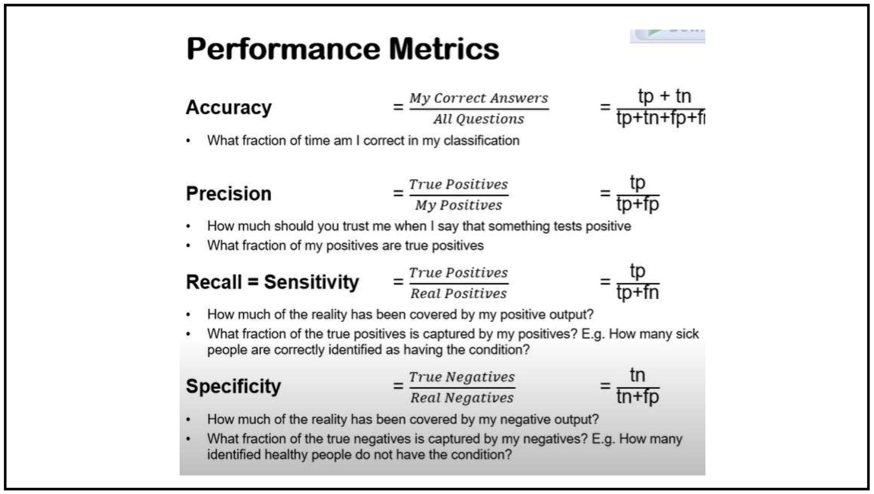
#	Height (inches)	Weight (kgs)	B.P. Sys	B.P. Dia	Heart disease				
	$\vec{x}$				y	$h(\vec{x})$			
1	62	70	120	80	No	No	True	Negative	
2	72	90	110	70	No	Yes	False	Positive	Type-I Error
3	74	80	130	70	No	No	True	Negative	
4	65	120	150	90	Yes	Yes	True	Positive	
5	67	100	140	85	Yes	No	False	Negative	Type-II Error
6	64	110	130	90	No	Yes	False	Positive	Type-I Error
7	69	150	170	100	Yes	Yes	True	Positive	
8	75	127	160	95	Yes	No	False	Negative	Type-II Error
9	66	66	135	90	Yes	Yes	True	Positive	

- y**:** Gold standards / Gold labels / Ground truth
- $h(\vec{x})$**:** Predicted labels
  - Positives and Negatives
  - True (matching the ground truth) or False (not matching the ground truth)

		gold standard labels	
		gold positive	gold negative
system output labels	system positive	**true positive**	**false positive**
	system negative	**false negative**	**true negative**



		Gold Labels	
		Gold Positive	Gold Negative
Predicted Labels	Predicted Positive	True Positives (tp)	False Positives (fp)
	Predicted Negative	False Negatives (fn)	True Negatives (tn)



Performance Metrics

Accuracy $= \frac{My\ Correct\ Answers}{All\ Questions} = \frac{tp + tn}{tp+tn+fp+f}$

- What fraction of time am I correct in my classification

Precision $= \frac{True\ Positives}{My\ Positives} = \frac{tp}{tp+fp}$

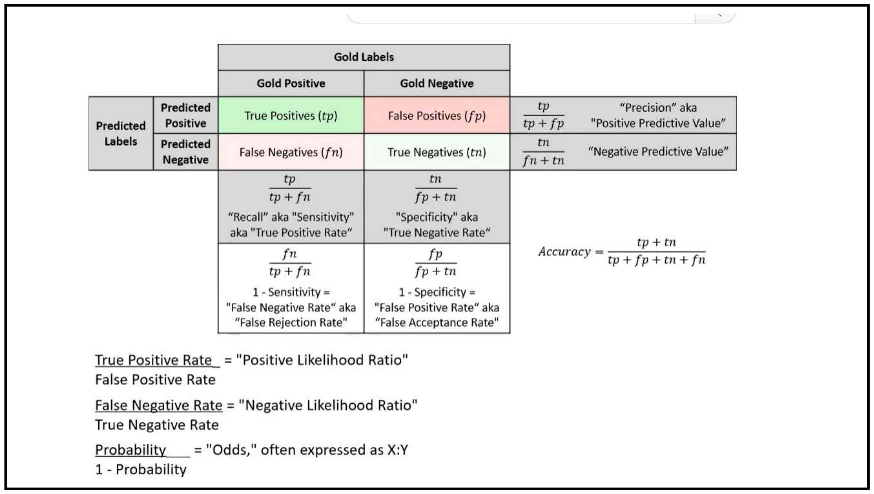
- How much should you trust me when I say that something tests positive
- What fraction of my positives are true positives

Recall = Sensitivity $= \frac{True\ Positives}{Real\ Positives} = \frac{tp}{tp+fn}$

- How much of the reality has been covered by my positive output?
- What fraction of the true positives is captured by my positives? E.g. How many sick people are correctly identified as having the condition?

Specificity $= \frac{True\ Negatives}{Real\ Negatives} = \frac{tn}{tn+fp}$

- How much of the reality has been covered by my negative output?
- What fraction of the true negatives is captured by my negatives? E.g. How many identified healthy people do not have the condition?



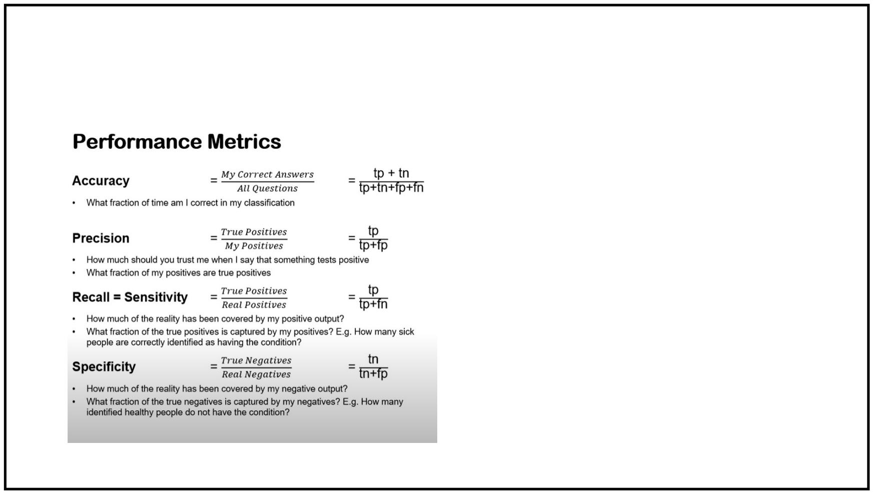
		Gold Labels			
		Gold Positive	Gold Negative		
Predicted Labels	Predicted Positive	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$	“Precision” aka "Positive Predictive Value”
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$	“Negative Predictive Value”
		$\frac{tp}{tp+fn}$ “Recall” aka "Sensitivity" aka "True Positive Rate“	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate“		
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate“ aka “False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate“ aka “False Acceptance Rate"		

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$

True Positive Rate = "Positive Likelihood Ratio"
False Positive Rate

False Negative Rate = "Negative Likelihood Ratio"
True Negative Rate

Probability = "Odds," often expressed as X:Y
1 - Probability



Performance Metrics

Accuracy $= \frac{My\ Correct\ Answers}{All\ Questions} = \frac{\text{tp + tn}}{\text{tp+tn+fp+fn}}$

- What fraction of time am I correct in my classification

Precision $= \frac{True\ Positives}{My\ Positives} = \frac{\text{tp}}{\text{tp+fp}}$

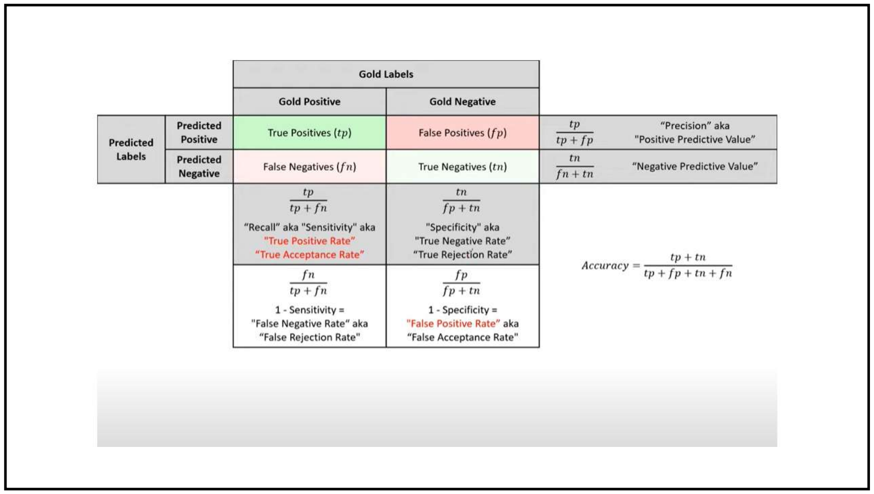
- How much should you trust me when I say that something tests positive
- What fraction of my positives are true positives

Recall = Sensitivity $= \frac{True\ Positives}{Real\ Positives} = \frac{\text{tp}}{\text{tp+fn}}$

- How much of the reality has been covered by my positive output?
- What fraction of the true positives is captured by my positives? E.g. How many sick people are correctly identified as having the condition?

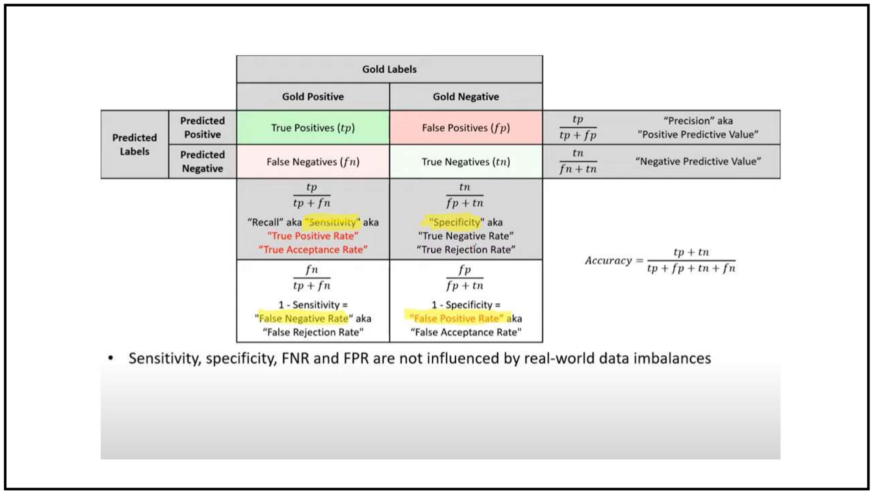
Specificity $= \frac{True\ Negatives}{Real\ Negatives} = \frac{\text{tn}}{\text{tn+fp}}$

- How much of the reality has been covered by my negative output?
- What fraction of the true negatives is captured by my negatives? E.g. How many identified healthy people do not have the condition?



		Gold Labels			
		Gold Positive	Gold Negative		
Predicted Labels	Predicted Positive	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$	“Precision” aka "Positive Predictive Value”
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$	“Negative Predictive Value”
		$\frac{tp}{tp+fn}$ “Recall” aka "Sensitivity" aka "True Positive Rate” “True Acceptance Rate”	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate” “True Rejection Rate”		
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate“ aka “False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate” aka “False Acceptance Rate"		

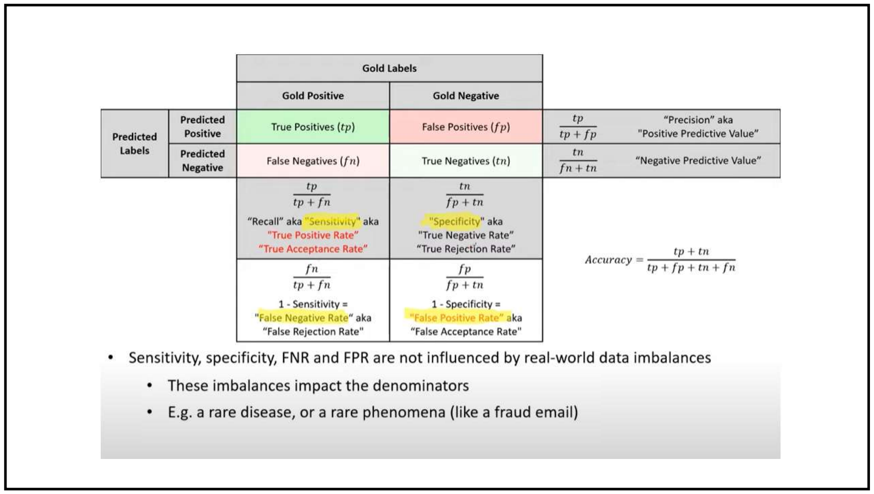
$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$



		Gold Labels			
		Gold Positive	Gold Negative		
Predicted Labels	Predicted Positive	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$	“Precision” aka "Positive Predictive Value”
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$	“Negative Predictive Value”
		$\frac{tp}{tp+fn}$ “Recall” aka "Sensitivity" aka "True Positive Rate” “True Acceptance Rate”	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate” “True Rejection Rate”		
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate“ aka “False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate” aka “False Acceptance Rate"		

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$

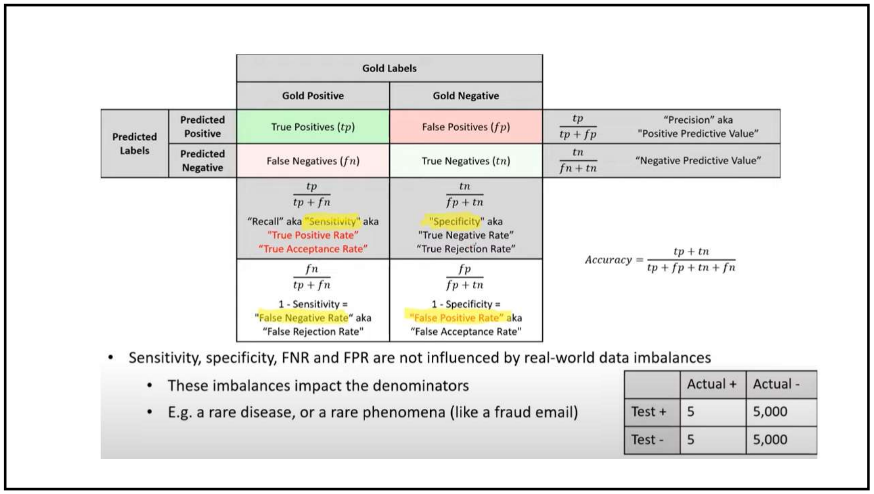
- Sensitivity, specificity, FNR and FPR are not influenced by real-world data imbalances



		Gold Labels		
		Gold Positive	Gold Negative	
Predicted Labels	**Predicted Positive**	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$ “Precision” aka "Positive Predictive Value”
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$ “Negative Predictive Value”
		$\frac{tp}{tp+fn}$ “Recall” aka "Sensitivity" aka "True Positive Rate” “True Acceptance Rate”	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate” “True Rejection Rate”	
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate“ aka “False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate” aka “False Acceptance Rate"	

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$

- Sensitivity, specificity, FNR and FPR are not influenced by real-world data imbalances
  - These imbalances impact the denominators
  - E.g. a rare disease, or a rare phenomena (like a fraud email)

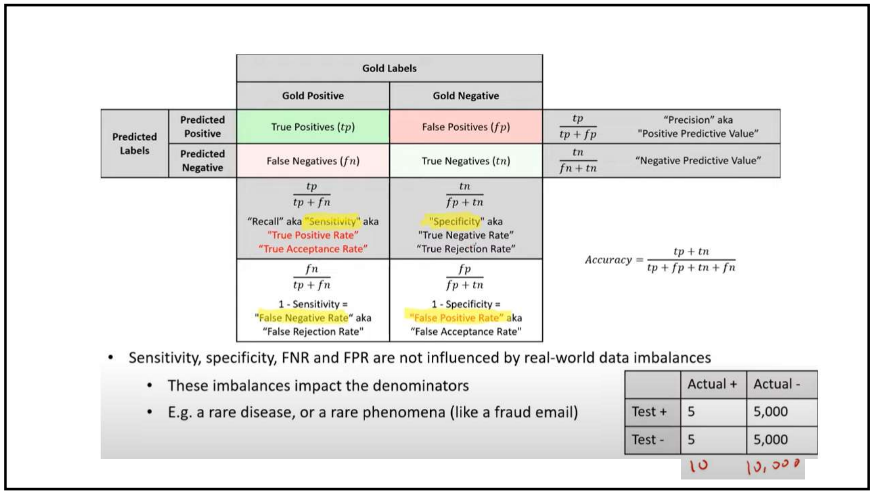


		Gold Labels		
		Gold Positive	Gold Negative	
Predicted Labels	**Predicted Positive**	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$ “Precision” aka "Positive Predictive Value”
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$ “Negative Predictive Value”
		$\frac{tp}{tp+fn}$ “Recall” aka "Sensitivity" aka "True Positive Rate” “True Acceptance Rate”	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate” “True Rejection Rate”	
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate“ aka “False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate” aka “False Acceptance Rate"	

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$

- Sensitivity, specificity, FNR and FPR are not influenced by real-world data imbalances
  - These imbalances impact the denominators
  - E.g. a rare disease, or a rare phenomena (like a fraud email)

	Actual +	Actual -
Test +	5	5,000
Test -	5	5,000

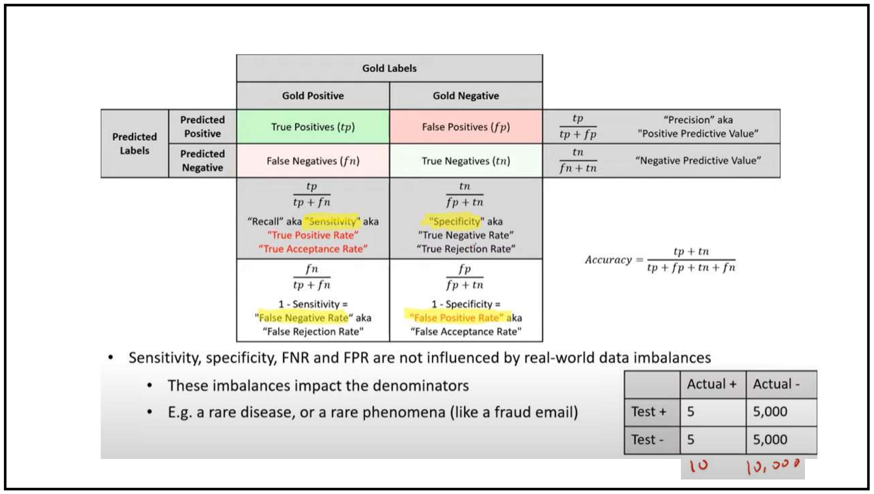


		Gold Labels			
		Gold Positive	**Gold Negative**		
Predicted Labels	**Predicted Positive**	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$	"Precision" aka "Positive Predictive Value"
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$	"Negative Predictive Value"
		$\frac{tp}{tp+fn}$ "Recall" aka "Sensitivity" aka "True Positive Rate" "True Acceptance Rate"	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate" "True Rejection Rate"		
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate" aka "False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate" aka "False Acceptance Rate"		

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

- Sensitivity, specificity, FNR and FPR are not influenced by real-world data imbalances
  - These imbalances impact the denominators
  - E.g. a rare disease, or a rare phenomena (like a fraud email)

	Actual +	Actual -
Test +	5	5,000
Test -	5	5,000
	10	10,000

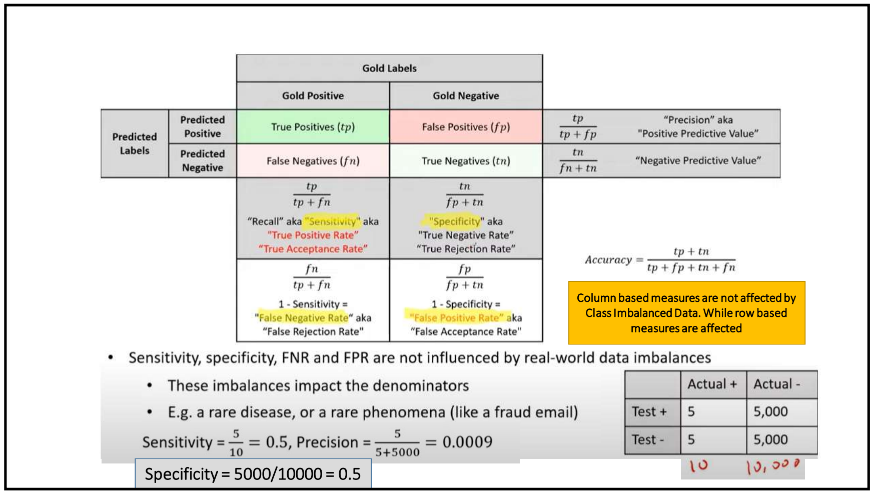


		Gold Labels			
		Gold Positive	**Gold Negative**		
Predicted Labels	**Predicted Positive**	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$	"Precision" aka "Positive Predictive Value"
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$	"Negative Predictive Value"
		$\frac{tp}{tp+fn}$ "Recall" aka "Sensitivity" aka "True Positive Rate" "True Acceptance Rate"	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate" "True Rejection Rate"		
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate" aka "False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate" aka "False Acceptance Rate"		

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

- Sensitivity, specificity, FNR and FPR are not influenced by real-world data imbalances
  - These imbalances impact the denominators
  - E.g. a rare disease, or a rare phenomena (like a fraud email)

	Actual +	Actual -
Test +	5	5,000
Test -	5	5,000
	10	10,000



		Gold Labels			
		Gold Positive	Gold Negative		
Predicted Labels	Predicted Positive	True Positives (tp)	False Positives (fp)	$\frac{tp}{tp+fp}$	“Precision” aka "Positive Predictive Value”
	Predicted Negative	False Negatives (fn)	True Negatives (tn)	$\frac{tn}{fn+tn}$	“Negative Predictive Value”
		$\frac{tp}{tp+fn}$ “Recall” aka "Sensitivity" aka "True Positive Rate” “True Acceptance Rate”	$\frac{tn}{fp+tn}$ "Specificity" aka "True Negative Rate” “True Rejection Rate”		
		$\frac{fn}{tp+fn}$ 1 - Sensitivity = "False Negative Rate“ aka “False Rejection Rate"	$\frac{fp}{fp+tn}$ 1 - Specificity = "False Positive Rate” aka “False Acceptance Rate"		

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

Column based measures are not affected by Class Imbalanced Data. While row based measures are affected

- Sensitivity, specificity, FNR and FPR are not influenced by real-world data imbalances
  - These imbalances impact the denominators
  - E.g. a rare disease, or a rare phenomena (like a fraud email)

Sensitivity = $\frac{5}{10} = 0.5$, Precision = $\frac{5}{5+5000} = 0.0009$

	Actual +	Actual -
Test +	5	5,000
Test -	5	5,000
	10	10,000

Specificity = 5000/10000 = 0.5

Receiver Operating Characteristic (RoC)

- A graphical plot that illustrates the diagnostic ability of a binary classifier as its discrimination threshold is varied

Receiver Operating Characteristic (RoC)

- A graphical plot that illustrates the diagnostic ability of a binary classifier as its discrimination threshold is varied
- The method was originally developed for operators of military radar receivers starting in 1941, which led to its name.

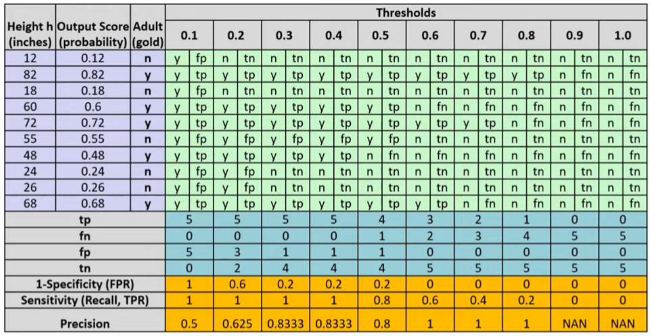
https://en.wikipedia.org/wiki/Receiver_operating_characteristic

---

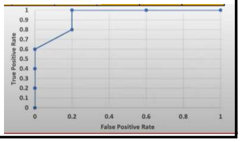
Receiver Operating Characteristic (RoC)

- A graphical plot that illustrates the diagnostic ability of a binary classifier as its discrimination threshold is varied
- The method was originally developed for operators of military radar receivers starting in 1941, which led to its name.
- Plot the *true positive rate* (TPR) – sensitivity – against the *false positive rate* (FPR) – (1 - specificity) at various threshold settings

https://en.wikipedia.org/wiki/Receiver_operating_characteristic

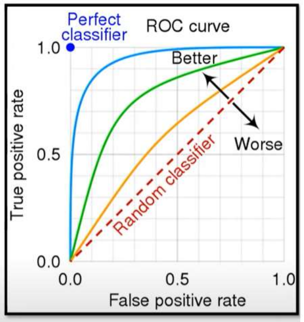


			Thresholds																			
Height h (inches)	Output Score (probability)	Adult (gold)	0.1		0.2		0.3		0.4		0.5		0.6		0.7		0.8		0.9		1.0	
12	0.12	n	y	fp	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn
82	0.82	y	y	tp	y	tp	y	tp	y	tp	y	tp	y	tp	y	tp	y	tp	n	fn	n	fn
18	0.18	n	y	fp	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn
60	0.6	y	y	tp	y	tp	y	tp	y	tp	y	tp	n	fn	n	fn	n	fn	n	fn	n	fn
72	0.72	y	y	tp	y	tp	y	tp	y	tp	y	tp	y	tp	y	tp	n	fn	n	fn	n	fn
55	0.55	n	y	fp	y	fp	y	fp	y	fp	y	fp	n	tn	n	tn	n	tn	n	tn	n	tn
48	0.48	y	y	tp	y	tp	y	tp	y	tp	n	fn	n	fn	n	fn	n	fn	n	fn	n	fn
24	0.24	n	y	fp	y	fp	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn
26	0.26	n	y	fp	y	fp	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn	n	tn
68	0.68	y	y	tp	y	tp	y	tp	y	tp	y	tp	y	tp	n	fn	n	fn	n	fn	n	fn
tp			5		5		5		5		4		3		2		1		0		0	
fn			0		0		0		0		1		2		3		4		5		5	
fp			5		3		1		1		1		0		0		0		0		0	
tn			0		2		4		4		4		5		5		5		5		5	
1-Specificity (FPR)			1		0.6		0.2		0.2		0.2		0		0		0		0		0	
Sensitivity (Recall, TPR)			1		1		1		1		0.8		0.6		0.4		0.2		0		0	
Precision			0.5		0.625		0.8333		0.8333		0.8		1		1		1		NAN		NAN	



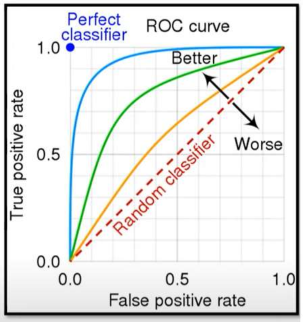
Characteristics

- The best possible prediction method would yield a point in the upper left corner (0,1) i.e.100% sensitivity (no false negatives) and 100% specificity (no false positives)



Characteristics

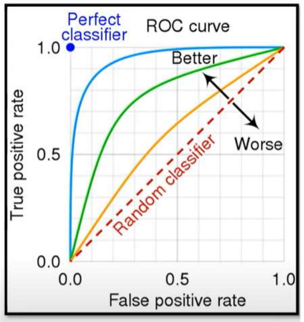
- The best possible prediction method would yield a point in the upper left corner (0,1) i.e.100% sensitivity (no false negatives) and 100% specificity (no false positives)
- A random guess would give a point along a diagonal line (the line of no-discrimination) from the left bottom to the top right corners (TPR = FPR)

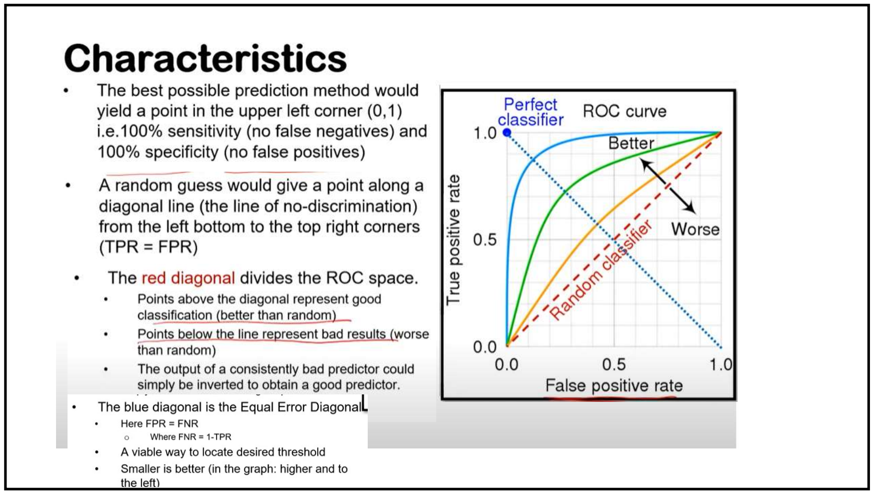


Characteristics

- The best possible prediction method would yield a point in the upper left corner (0,1) i.e.100% sensitivity (no false negatives) and 100% specificity (no false positives)
- A random guess would give a point along a diagonal line (the line of no-discrimination) from the left bottom to the top right corners (TPR = FPR)

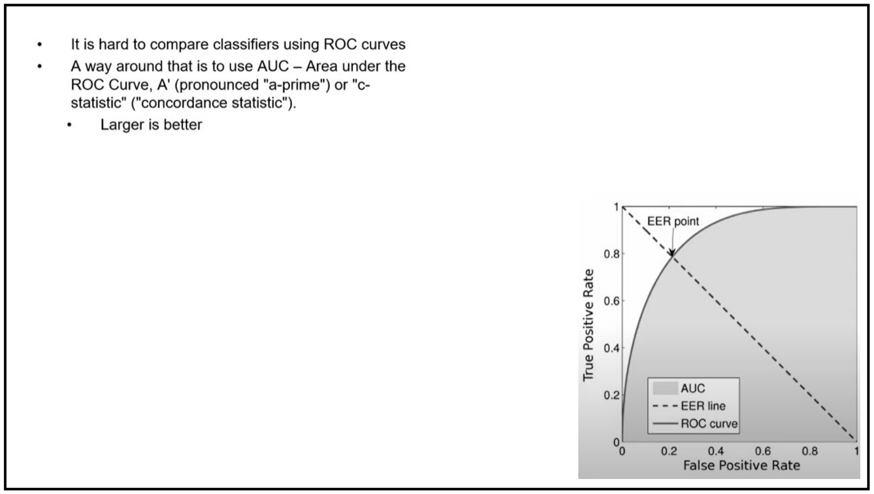
- The red diagonal divides the ROC space.
  - Points above the diagonal represent good classification (better than random)
  - Points below the line represent bad results (worse than random)
  - The output of a consistently bad predictor could simply be inverted to obtain a good predictor.





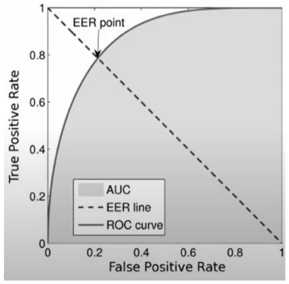
Characteristics

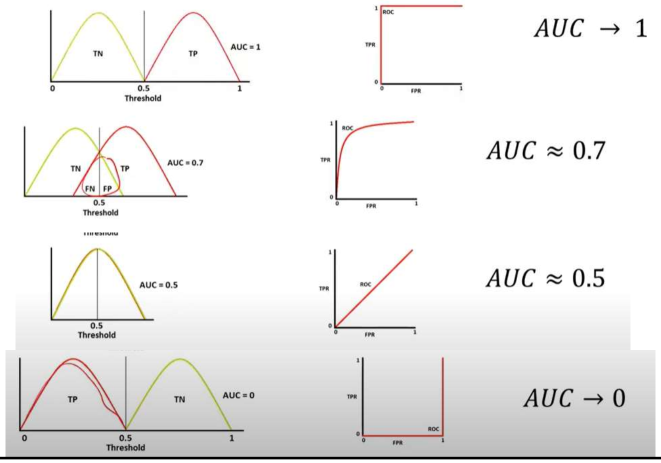
- The best possible prediction method would yield a point in the upper left corner (0,1) i.e.100% sensitivity (no false negatives) and 100% specificity (no false positives)
- A random guess would give a point along a diagonal line (the line of no-discrimination) from the left bottom to the top right corners (TPR = FPR)
- The red diagonal divides the ROC space.
  - Points above the diagonal represent good classification (better than random)
  - Points below the line represent bad results (worse than random)
  - The output of a consistently bad predictor could simply be inverted to obtain a good predictor.
- The blue diagonal is the Equal Error Diagonal
  - Here FPR = FNR
    - Where FNR = 1-TPR
  - A viable way to locate desired threshold
  - Smaller is better (in the graph: higher and to the left)



- It is hard to compare classifiers using ROC curves
- A way around that is to use AUC – Area under the ROC Curve, A' (pronounced "a-prime") or "c-statistic" ("concordance statistic").
  - Larger is better

- It is hard to compare classifiers using ROC curves
- A way around that is to use AUC – Area under the ROC Curve, A' (pronounced "a-prime") or "c-statistic" ("concordance statistic").
  - Larger is better
- "AUC ROC can be interpreted as the probability that the scores given by a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one." (Page 54, Learning from Imbalanced Data Sets, 2018)
- For imbalanced datasets: "ROC analysis does not have any bias toward models that perform well on the minority class at the expense of the majority class—a property that is quite attractive when dealing with imbalanced data." (Page 27, Imbalanced Learning: Foundations, Algorithms, and Applications, 2013)





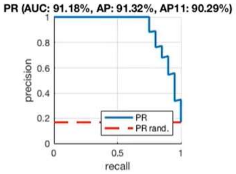
$AUC \rightarrow 1$

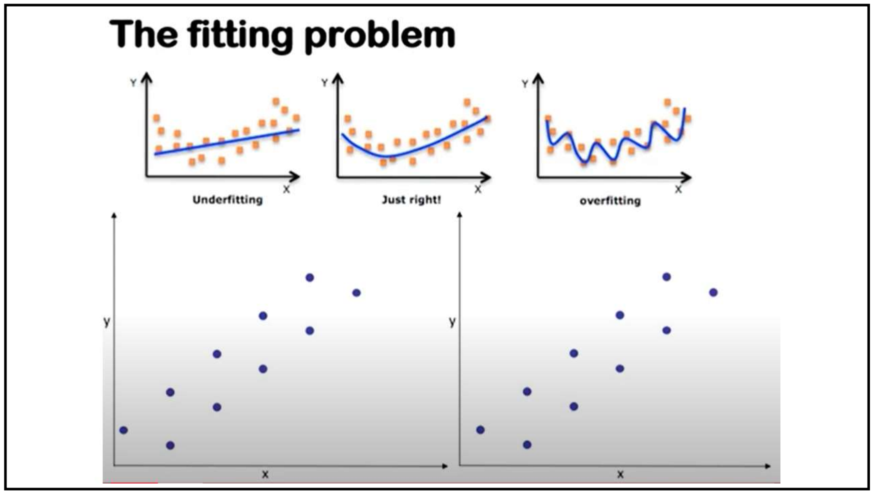
$AUC \approx 0.7$

$AUC \approx 0.5$

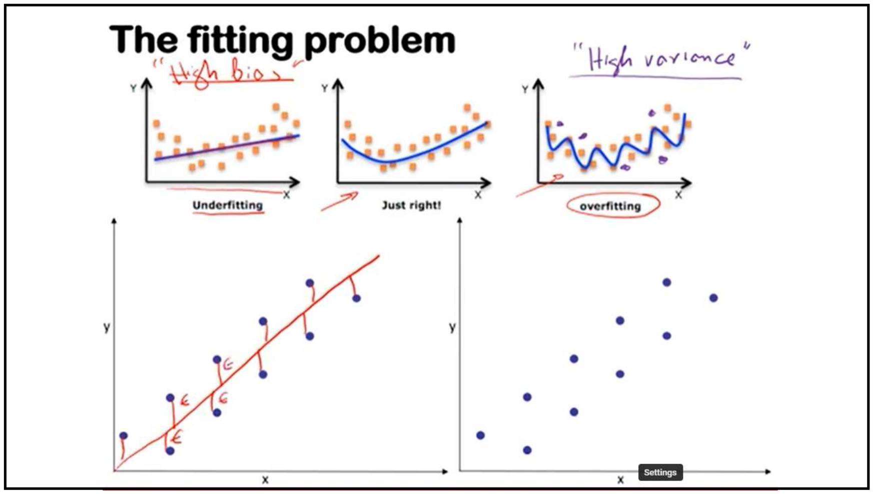
$AUC \rightarrow 0$

- The precision-recall curve shows the tradeoff between precision and recall for different thresholds. A high area under the curve represents both high recall and high precision
- High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).





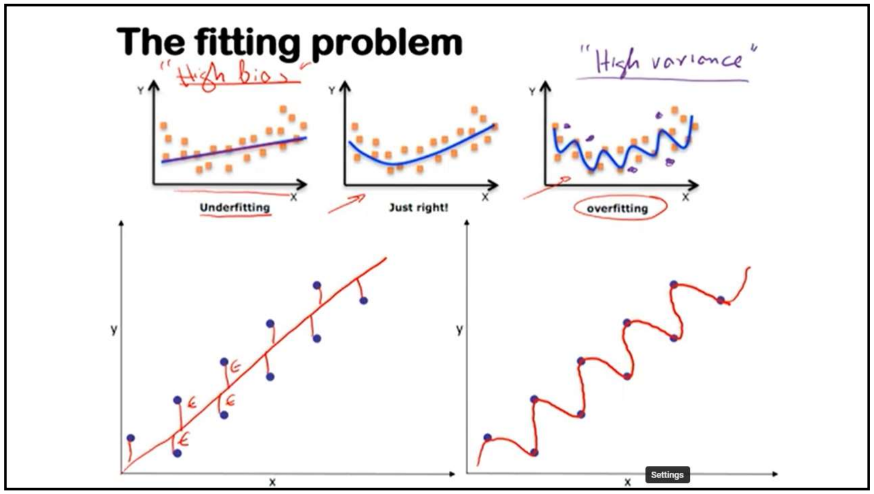
The fitting problem



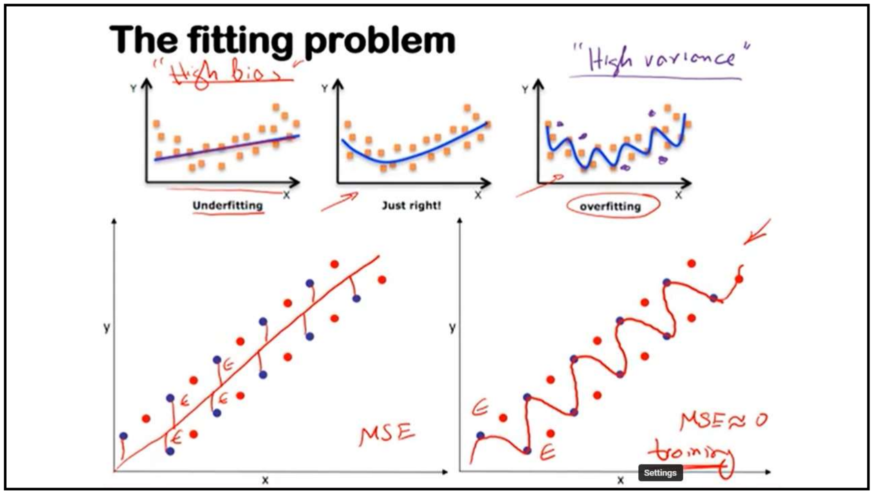
The fitting problem

"High bias"

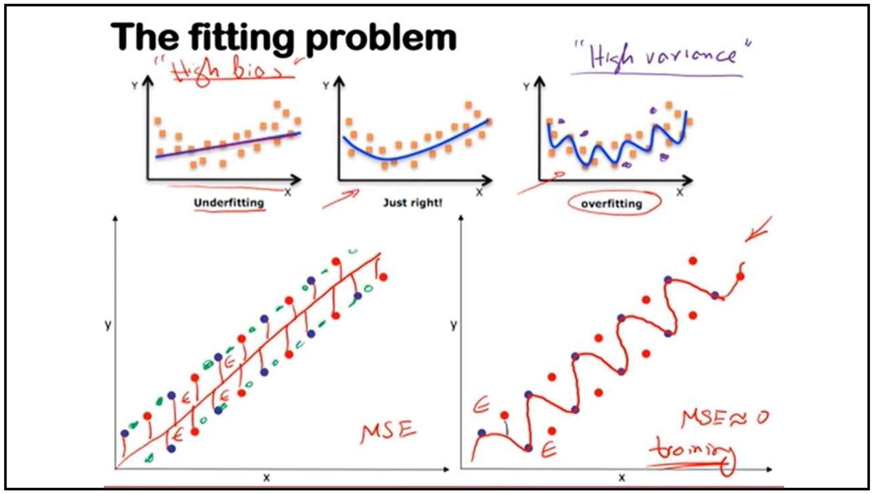
"High variance"



The fitting problem



The fitting problem



The fitting problem

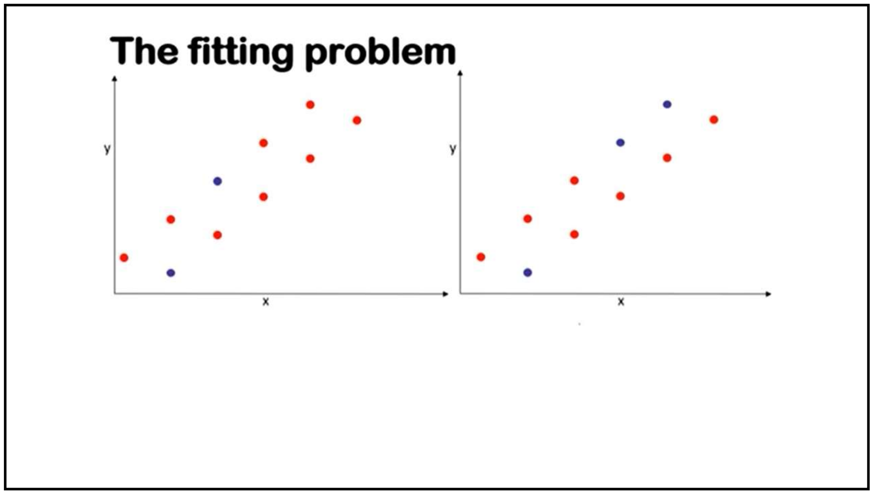
A real example

http://www.petanimalexpo.co.nz/

A real example

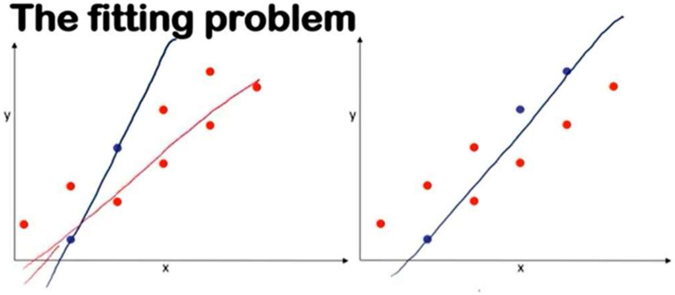
only occur
in the
→
training
Data
nois

http://www.petanimalexpo.co.nz/

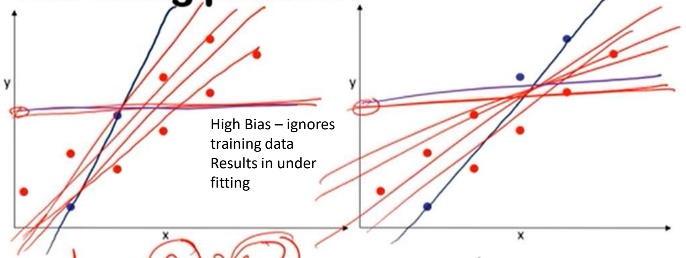


The fitting problem

The fitting problem

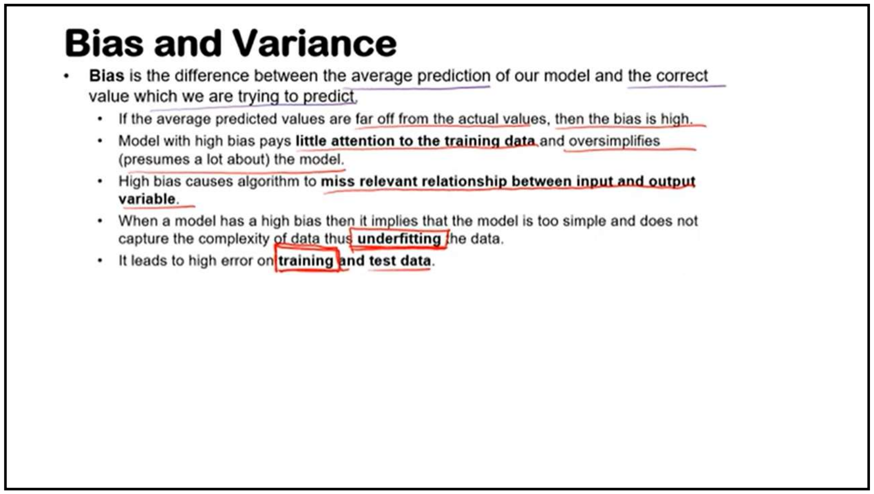


The fitting problem



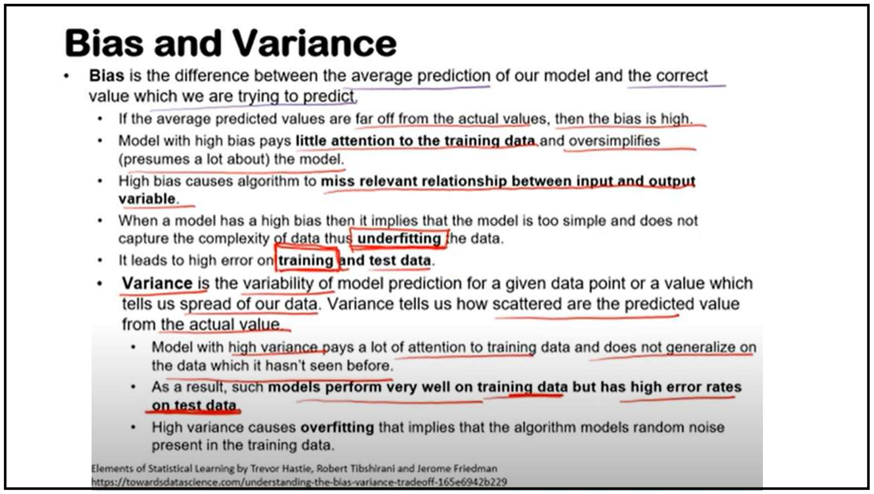
High Bias – ignores training data
Results in under fitting

$h_\theta(x) = \boxed{\theta_0} + \theta_1 \boxed{x_1}$



Bias and Variance

- **Bias** is the difference between the average prediction of our model and the correct value which we are trying to predict.
  - If the average predicted values are far off from the actual values, then the bias is high.
  - Model with high bias pays **little attention to the training data** and oversimplifies (presumes a lot about) the model.
  - High bias causes algorithm to **miss relevant relationship between input and output variable**.
  - When a model has a high bias then it implies that the model is too simple and does not capture the complexity of data thus **underfitting** the data.
  - It leads to high error on **training and test data**.

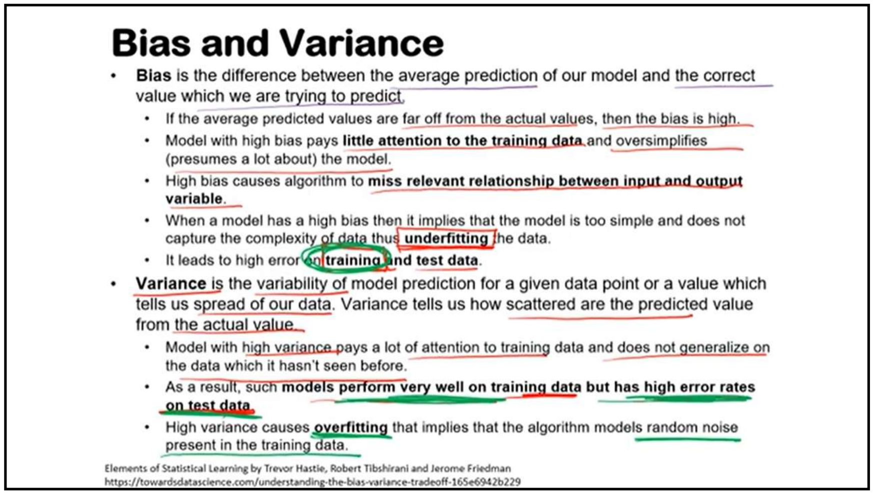


Bias and Variance

- **Bias** is the difference between the average prediction of our model and the correct value which we are trying to predict.
  - If the average predicted values are far off from the actual values, then the bias is high.
  - Model with high bias pays **little attention to the training data** and oversimplifies (presumes a lot about) the model.
  - High bias causes algorithm to **miss relevant relationship between input and output variable**.
  - When a model has a high bias then it implies that the model is too simple and does not capture the complexity of data thus **underfitting** the data.
  - It leads to high error on **training and test data**.

  - **Variance** is the variability of model prediction for a given data point or a value which tells us spread of our data. Variance tells us how scattered are the predicted value from the actual value.
    - Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before.
    - As a result, such **models perform very well on training data but has high error rates on test data**.
    - High variance causes **overfitting** that implies that the algorithm models random noise present in the training data.

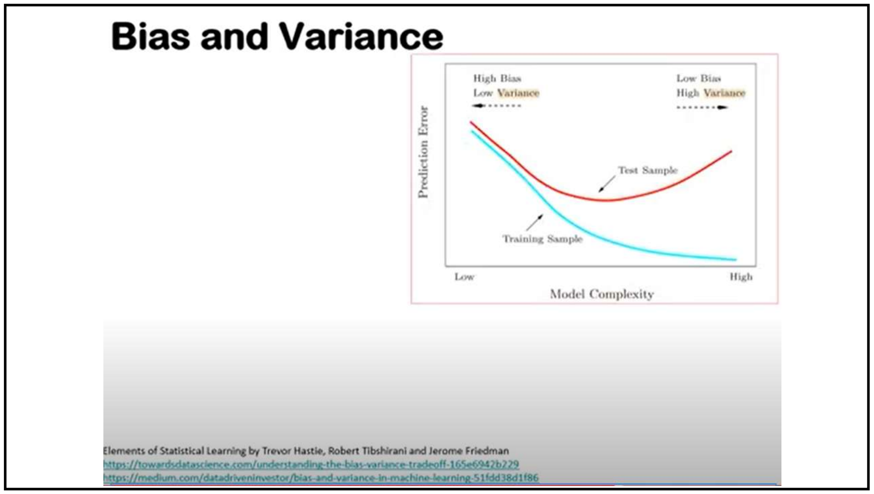
Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229



Bias and Variance

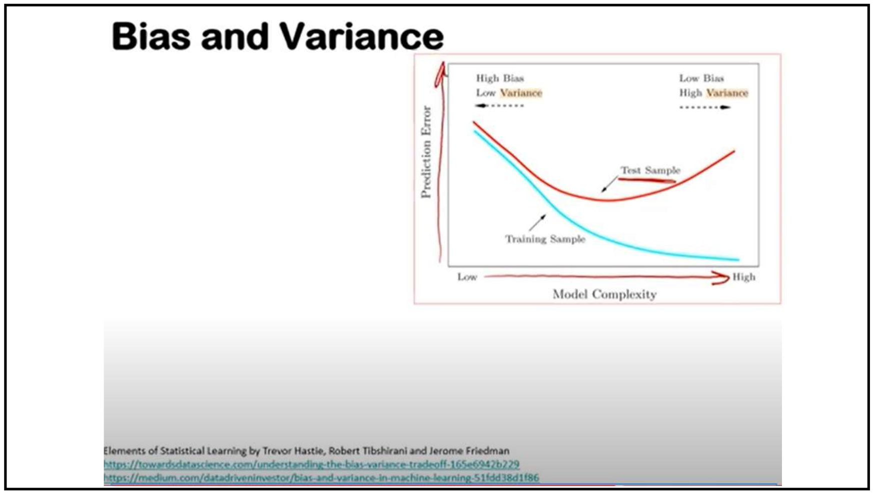
- **Bias** is the difference between the average prediction of our model and the correct value which we are trying to predict.
  - If the average predicted values are far off from the actual values, then the bias is high.
  - Model with high bias pays **little attention to the training data** and oversimplifies (presumes a lot about) the model.
  - High bias causes algorithm to **miss relevant relationship between input and output variable**.
  - When a model has a high bias then it implies that the model is too simple and does not capture the complexity of data thus **underfitting** the data.
  - It leads to high error on **training and test data**.
- **Variance** is the variability of model prediction for a given data point or a value which tells us spread of our data. Variance tells us how scattered are the predicted value from the actual value.
  - Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before.
  - As a result, such **models perform very well on training data but has high error rates on test data**.
  - High variance causes **overfitting** that implies that the algorithm models random noise present in the training data.

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229



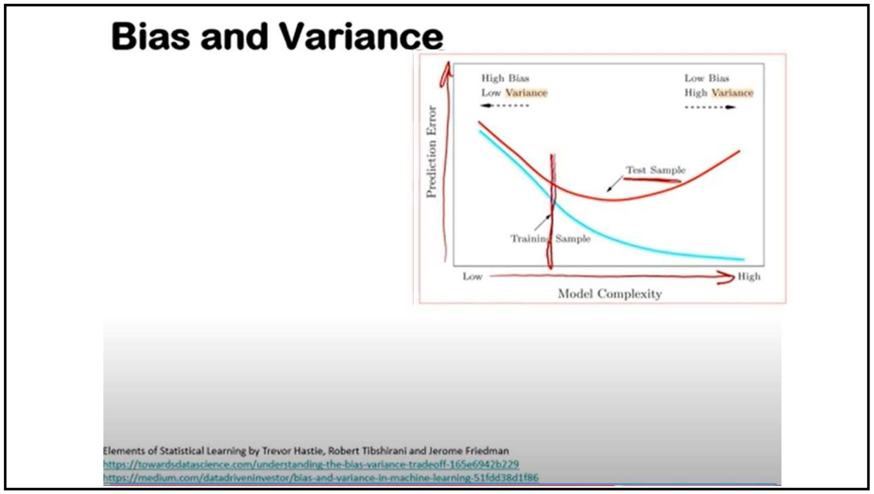
Bias and Variance

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



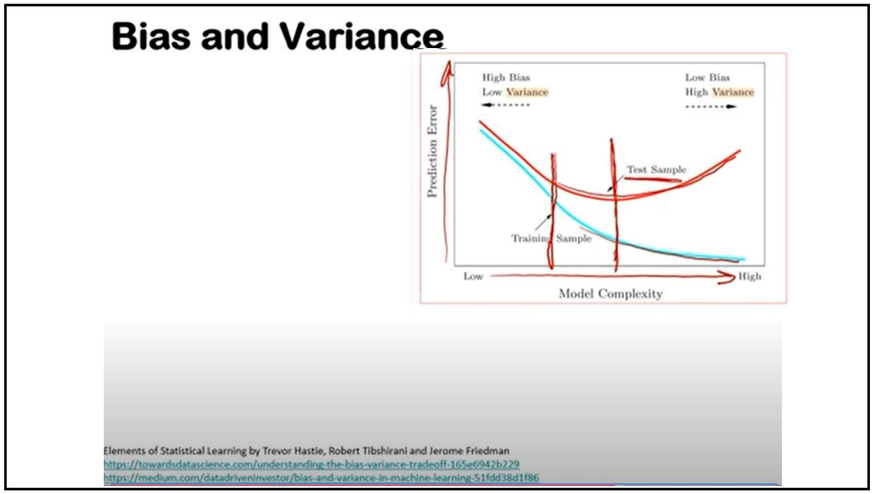
Bias and Variance

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



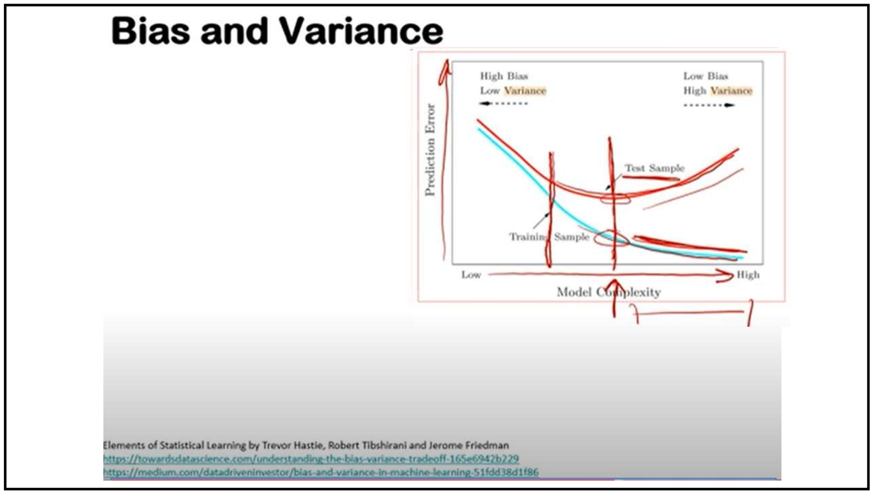
Bias and Variance

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



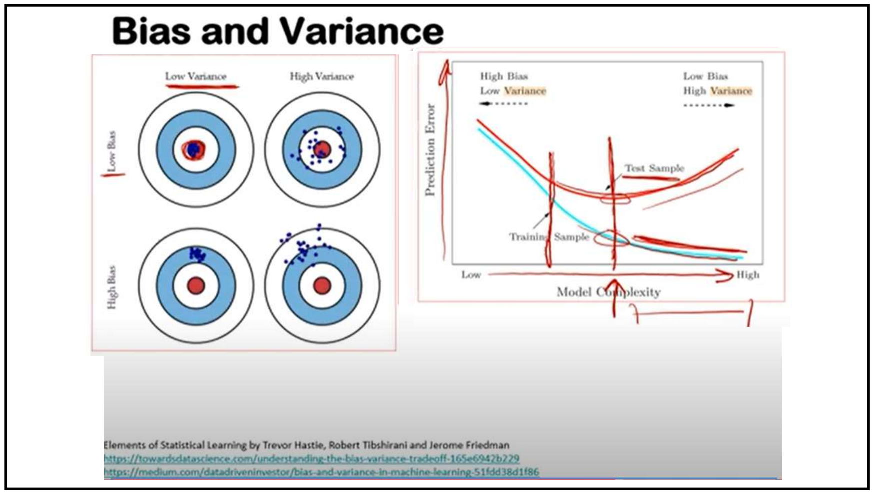
Bias and Variance

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



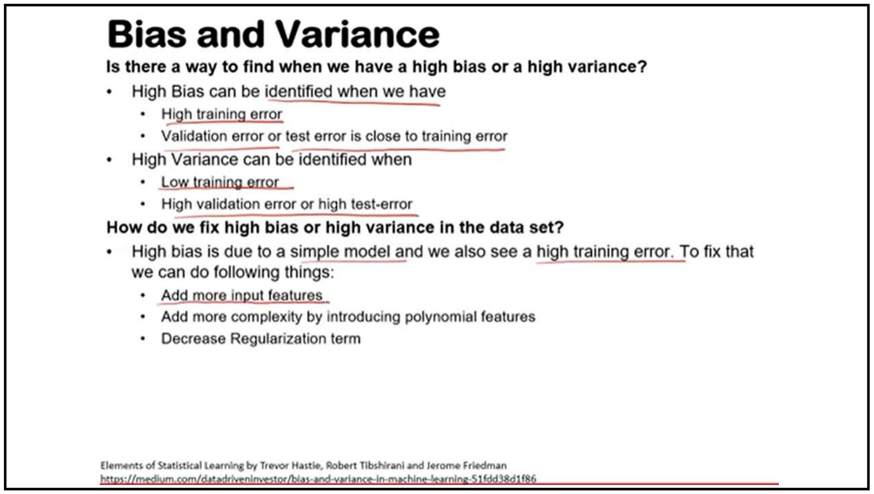
Bias and Variance

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



Bias and Variance

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



Bias and Variance

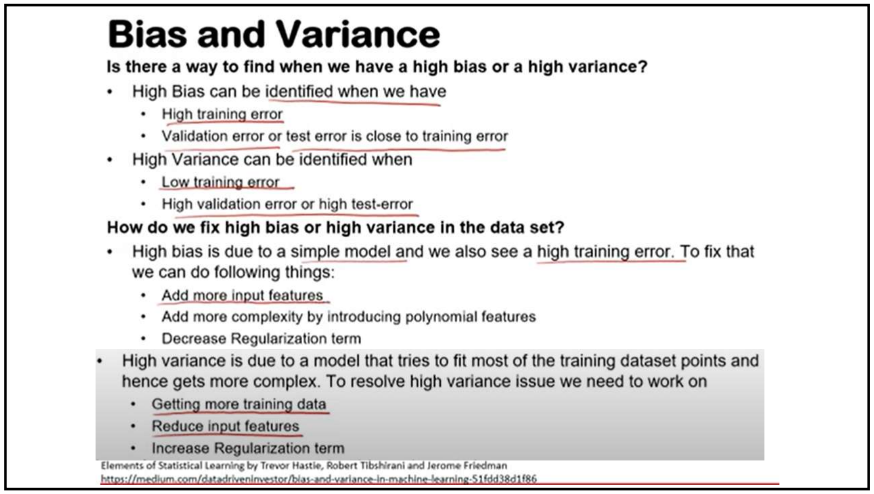
Is there a way to find when we have a high bias or a high variance?

- High Bias can be identified when we have
  - High training error
  - Validation error or test error is close to training error
- High Variance can be identified when
  - Low training error
  - High validation error or high test-error

How do we fix high bias or high variance in the data set?

- High bias is due to a simple model and we also see a high training error. To fix that we can do following things:
  - Add more input features
  - Add more complexity by introducing polynomial features
  - Decrease Regularization term

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



Bias and Variance

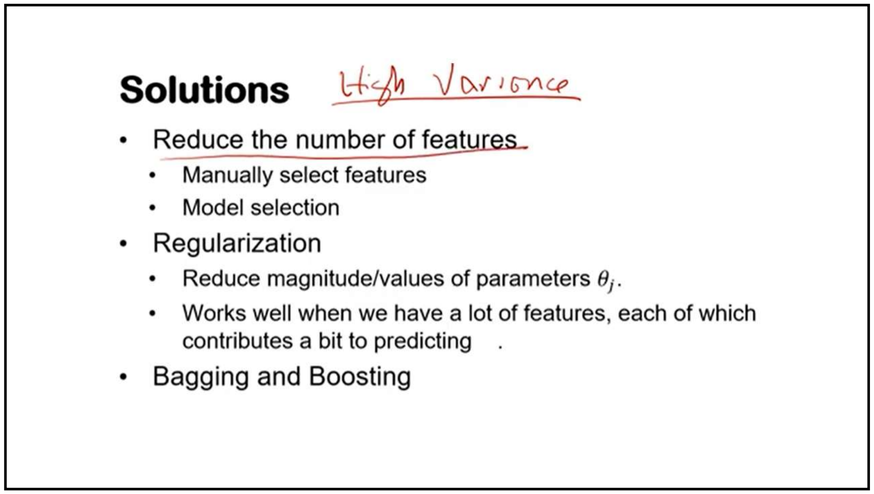
Is there a way to find when we have a high bias or a high variance?

- High Bias can be identified when we have
  - High training error
  - Validation error or test error is close to training error
- High Variance can be identified when
  - Low training error
  - High validation error or high test-error

How do we fix high bias or high variance in the data set?

- High bias is due to a simple model and we also see a high training error. To fix that we can do following things:
  - Add more input features
  - Add more complexity by introducing polynomial features
  - Decrease Regularization term
- High variance is due to a model that tries to fit most of the training dataset points and hence gets more complex. To resolve high variance issue we need to work on
  - Getting more training data
  - Reduce input features
  - Increase Regularization term

Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani and Jerome Friedman
https://medium.com/datadriveninvestor/bias-and-variance-in-machine-learning-51fdd38d1f86



Solutions

High Variance

- Reduce the number of features
  - Manually select features
  - Model selection
- Regularization
  - Reduce magnitude/values of parameters θ_j.
  - Works well when we have a lot of features, each of which contributes a bit to predicting .
- Bagging and Boosting