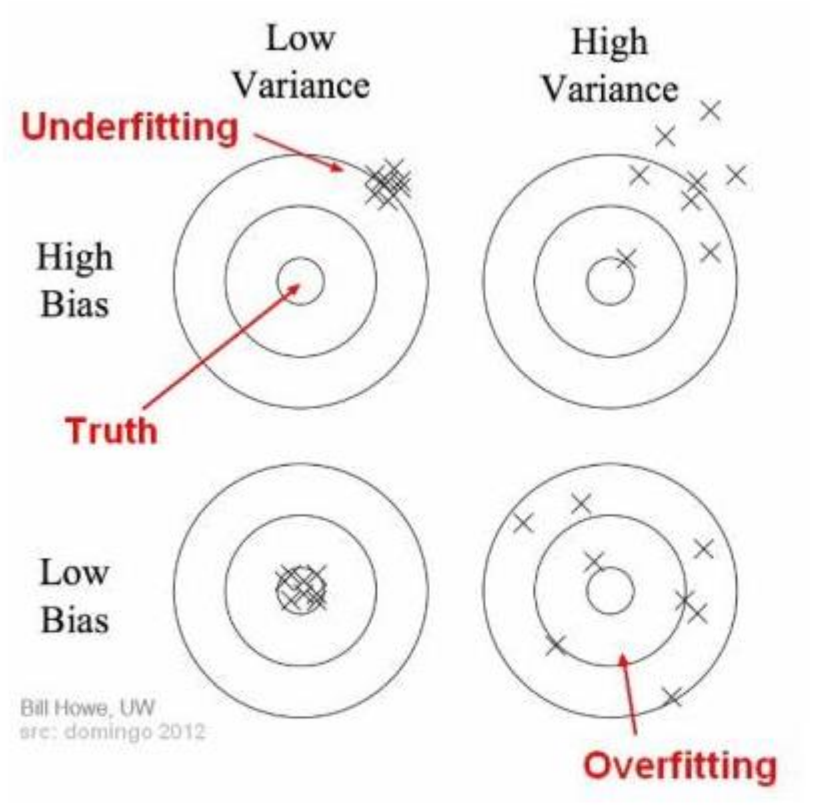
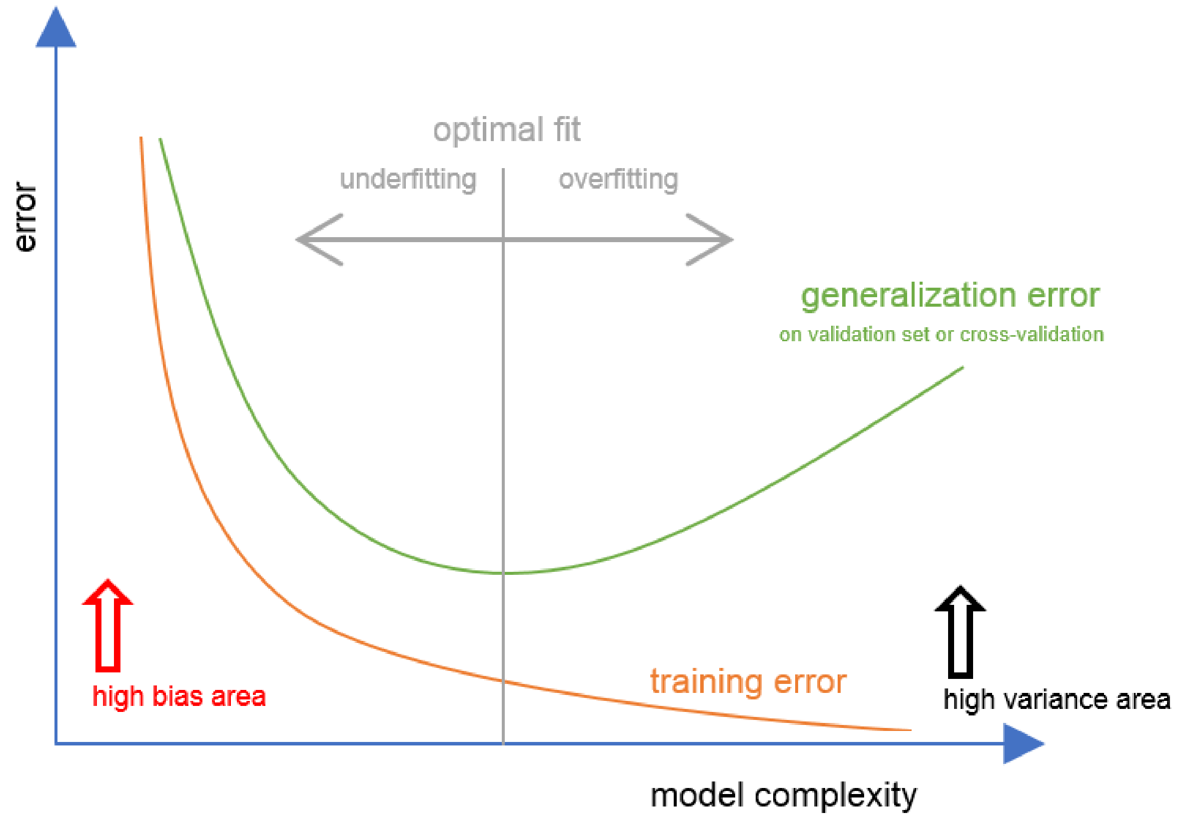
Bias Variance Tradeoff

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict.

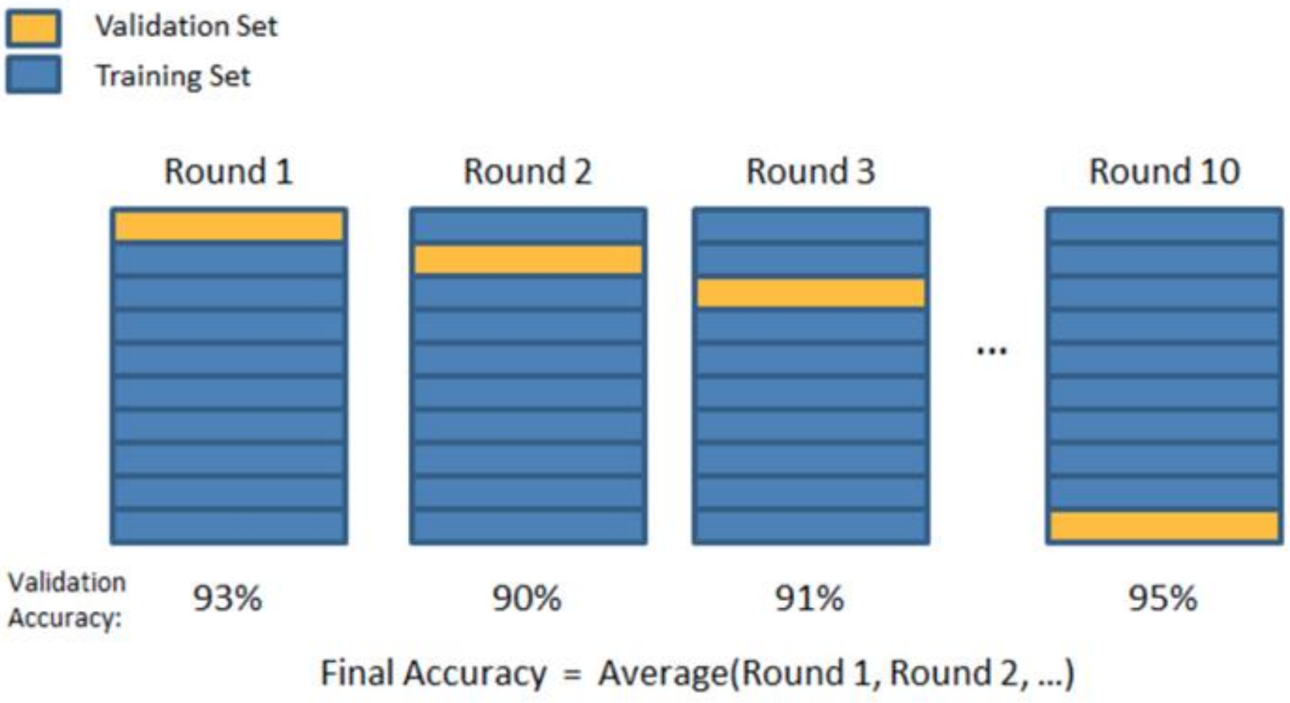
Variance is the variability of model prediction for a given data point or a value which tells us spread of our data.





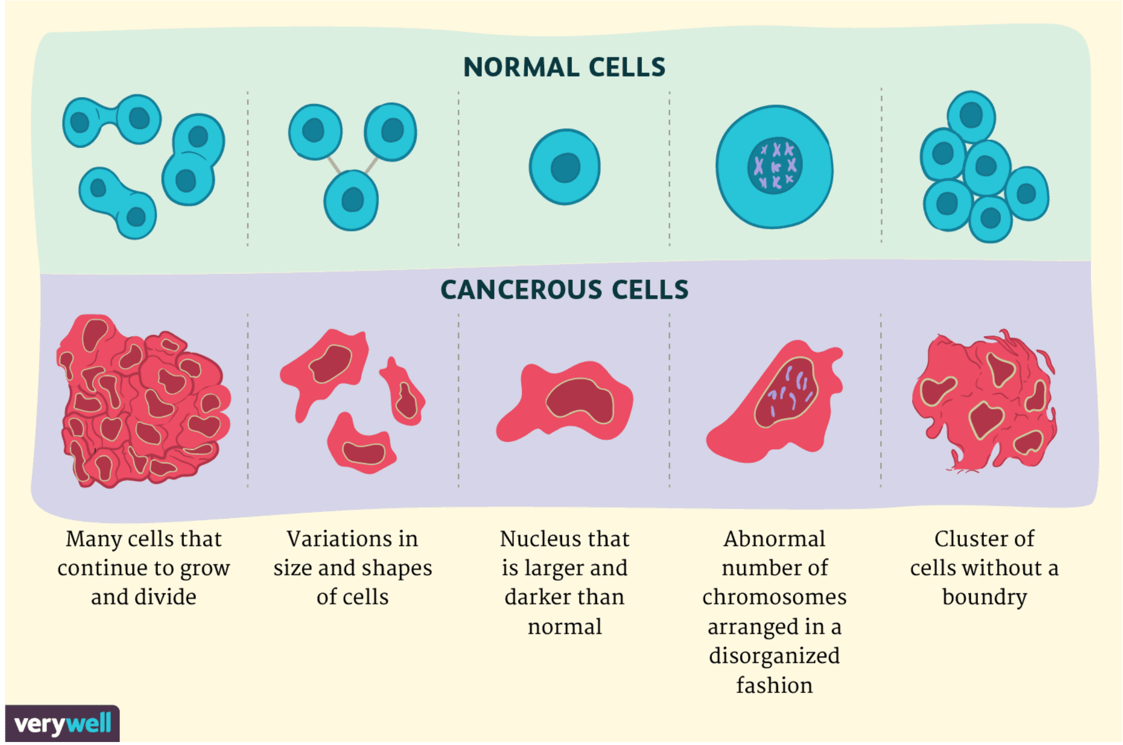
K-fold cross validation

- In K-Folds Cross Validation we split our data into k different subsets (or folds).
- We use k-1 subsets to train our data and leave the last subset (or the last fold) as test data.
- We then average the model against each of the folds and then finalize our model. After that we test it against the test set.



https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6

Evaluation Matrices



Normal cells: 90, Cancerous cells :10

Normal cells correctly identified : 90

Cancerous cells correctly identified : 0

Accuracy of the classifier : 90%

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

$$Precision = \frac{0}{0+1} = 0 \qquad Recall = \frac{0}{0+10} = 0$$

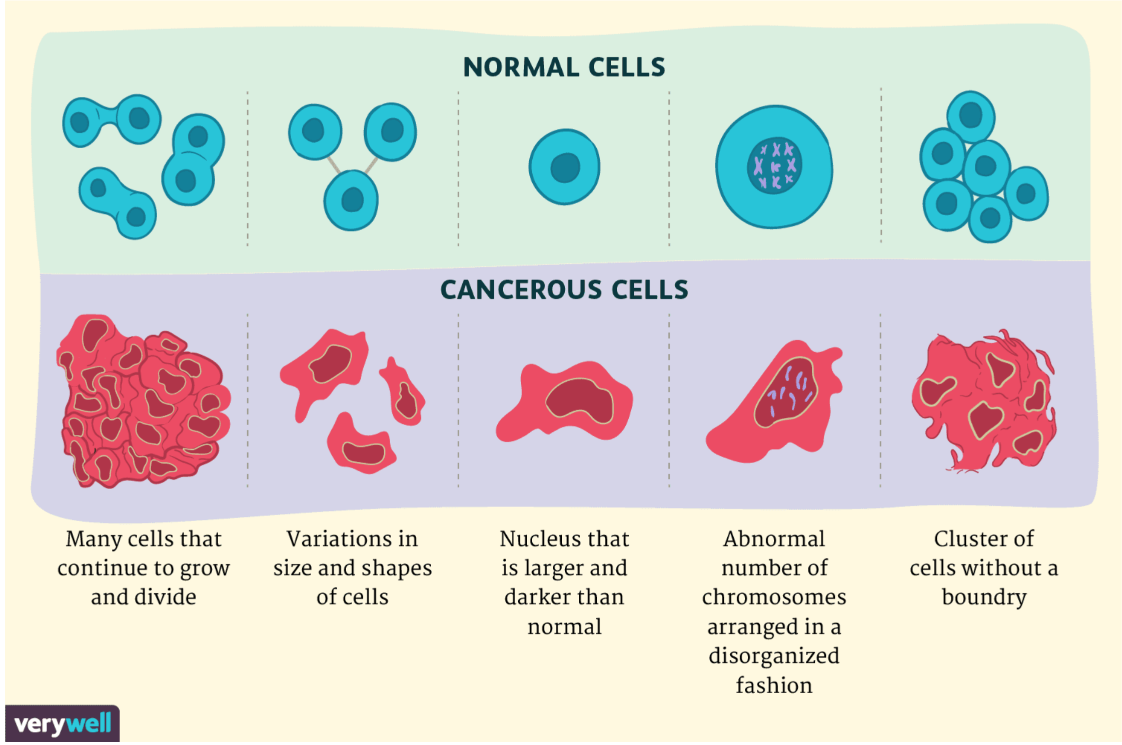
Predicted / Classified

Actual		**Normal**	**Cancerous**
	Normal	90	0
	Cancerous	10	0

Predicted

Actual		**Negative**	**Positive**
	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

confusion matrix is a table that is often used to **describe the performance of a classification model**



Normal cells: 50, Cancerous cells :50

Normal cells correctly identified : 30

Cancerous cells correctly identified : 20

Accuracy of the classifier : 50%

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Predicted / Classified

Actual		**Normal**	**Cancerous**
	Normal	30	20
	Cancerous	30	20

$$Precision = \frac{20}{20+20} = 0.5$$

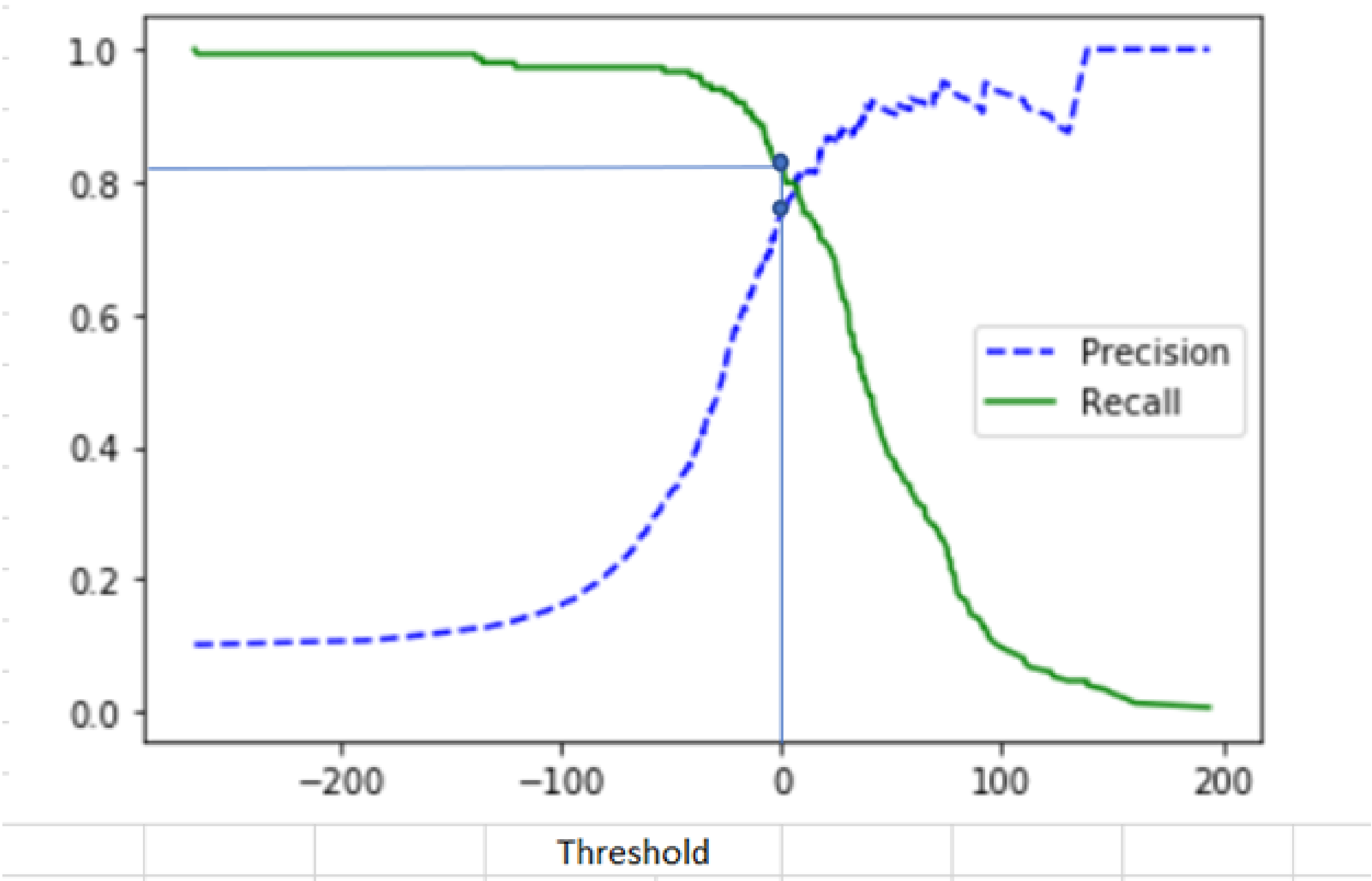
Predicted

Actual		**Negative**	**Positive**
	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

$$Precision = \frac{20}{20+20} = 0.5 \qquad Recall = \frac{20}{20+30} = 0.4$$

Precision/Recall Tradeoff

Unfortunately, when we try to improve precision, recall may come down, and vice versa. This is called the *precision/recall tradeoff*.



https://medium.com/analytics-vidhya/precision-recall-tradeoff-79e892d43134

Confusion Matrix for Multi-Class Classification

Predicted Class \ True Class	Apple	Orange	Mango
Apple	7	8	9
Orange	1	2	3
Mango	3	2	1

Class	Precision	Recall	F1-score
Apple	0.29	0.64	0.40
Orange	0.33	0.17	0.22
Mango	0.17	0.08	0.11

- ***TP = 7***
- ***TN = (2+3+2+1) = 8***
- ***FP = (8+9) = 17***
- ***FN = (1+3) = 4***

- ***Precision = 7/(7+17) = 0.29***
- ***Recall = 7/(7+4) = 0.64***
- ***F1-score = 0.40***

Unlike binary classification, there are no positive or negative classes here.

What we have to do here is to find TP, TN, FP and FN for each individual class.

Source : https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826

Predicted Class \ True Class	Apple	Orange	Mango
Apple	7	8	9
Orange	1	2	3
Mango	3	2	1

Micro F1

This is called micro-averaged F1-score. It is calculated by considering the total TP, total FP and total FN of the model.

•*Total TP = (7+2+1) = 10*
•*Total FP = (8+9)+(1+3)+(3+2) = 26*
•*Total FN = (1+3)+(8+2)+(9+3) = 26*

•***Precision = 10/(10+26) = 0.28***
•***Recall = 10/(10+26) = 0.28***

Micro F1 = 0.28

We can also combine the F1-score of each class to have a single measure for the whole model.

Class	Precision	Recall	F1-score
Apple	0.29	0.64	0.40
Orange	0.33	0.17	0.22
Mango	0.17	0.08	0.11

Macro F1

•*Class Apple F1-score = 0.40*
•*Class Orange F1-score = 0.22*
•*Class Mango F1-score = 0.11*

Macro F1 = (0.40+0.22+0.11)/3 = 0.24

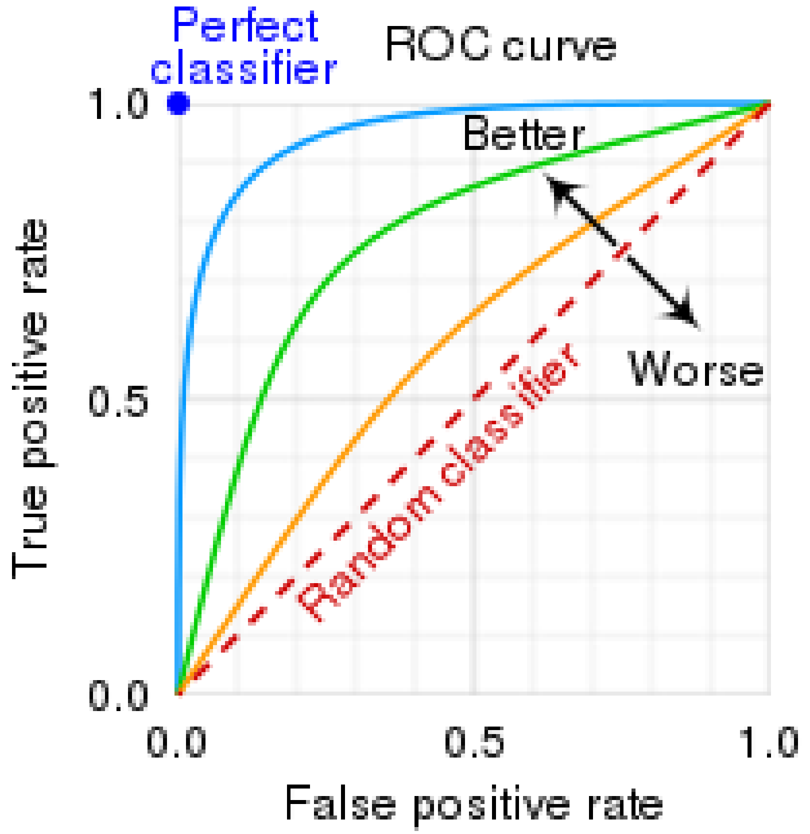
Weighted F1

It takes a weighted mean of the measures.

Weighted F1 =
((0.40*11)+(0.22*12)+(0.11*13))/(11+12+13) = 0.24

ROC curve

A **receiver operating characteristic curve**, or **ROC curve**, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.



$$\text{TPR /Recall / Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{FPR} = 1 - \text{Specificity}$$

$$= \frac{FP}{TN + FP}$$

AUC stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1).

AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.