Model Evaluation

Receiver Operating Characteristic (ROC) Curves and Area Under the Curve (AUC)

Overview

- Receiver Operating Characteristic (ROC) Curves
- Precision Recall (PR) Curves
- Comparing models using ROC AUC / PR AUC

Alternate Accuracy Measures

		Actual Class	
		C_1	C_2
Predicted Class	C_1	**$n_{1,1}$ = number of C_1 records classified correctly as C_1** **True Positive (TP)**	**$n_{2,1}$ = number of C_2 records classified incorrectly as C_1** **False Positive (FP)**
	C_2	**$n_{1,2}$ = number of C_1 records classified incorrectly as C_2** **False Negative (FN)**	**$n_{2,2}$ = number of C_2 records classified correctly as C_2** **True Negative (TN)**

- If "C_1" is the important class,
- **Sensitivity (also called "recall)** = % of actual C_1 class correctly classified

$$n_{1,1} / (n_{1,1} + n_{1,2})$$

True Positive Rate, TPR = TP/ (TP + FN)

- **Specificity** = % of actual C_2 class correctly classified

$$n_{2,2} / (n_{2,1} + n_{2,2})$$

True Negative Rate, TNR = TN / (FP + TN)

- **False Positive Rate (FPR)** = 1- Specificity

FPR = FP / (FP + TN)

ROC Curves (Receiver Operating Characteristic Curves) for a Perfect Classifier

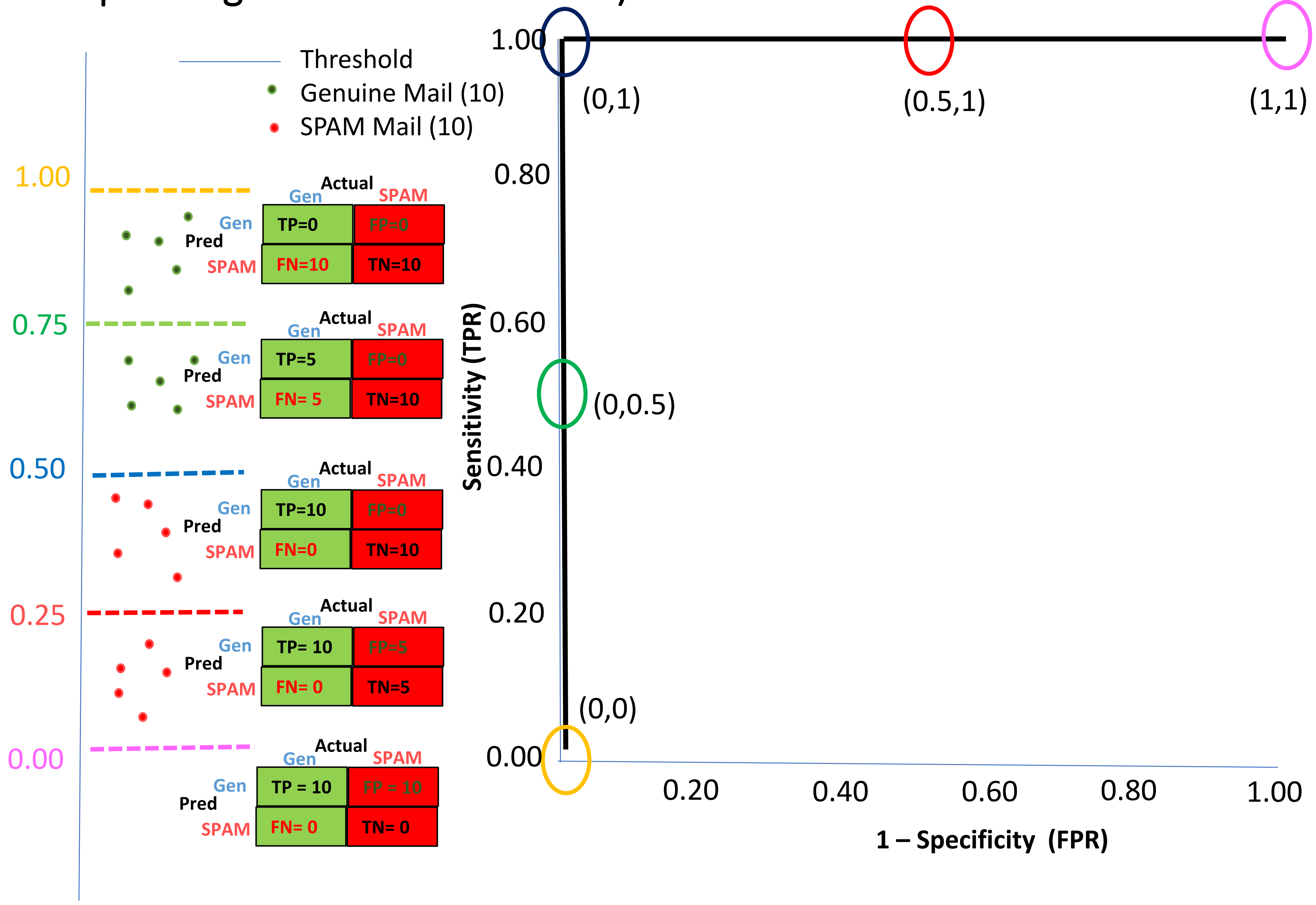
When threshold is near 1,
Sensitivity = TP/(TP+FN) = 0/10 = 0
Specificity = TN/(TN+FP) = 10/10= 1
Hence (0,0)

When threshold is near .75,
Sensitivity = TP/(TP+FN) = 5/10 = 0.5
Specificity = TN/(TN+FP) = 10/10= 1
Hence (0,0.5)

When threshold is near 0.5,
Sensitivity = TP/(TP+FN) = 10/10 = 1
Specificity = TN/(TN+FP) = 10/10= 1
Hence (0,1)

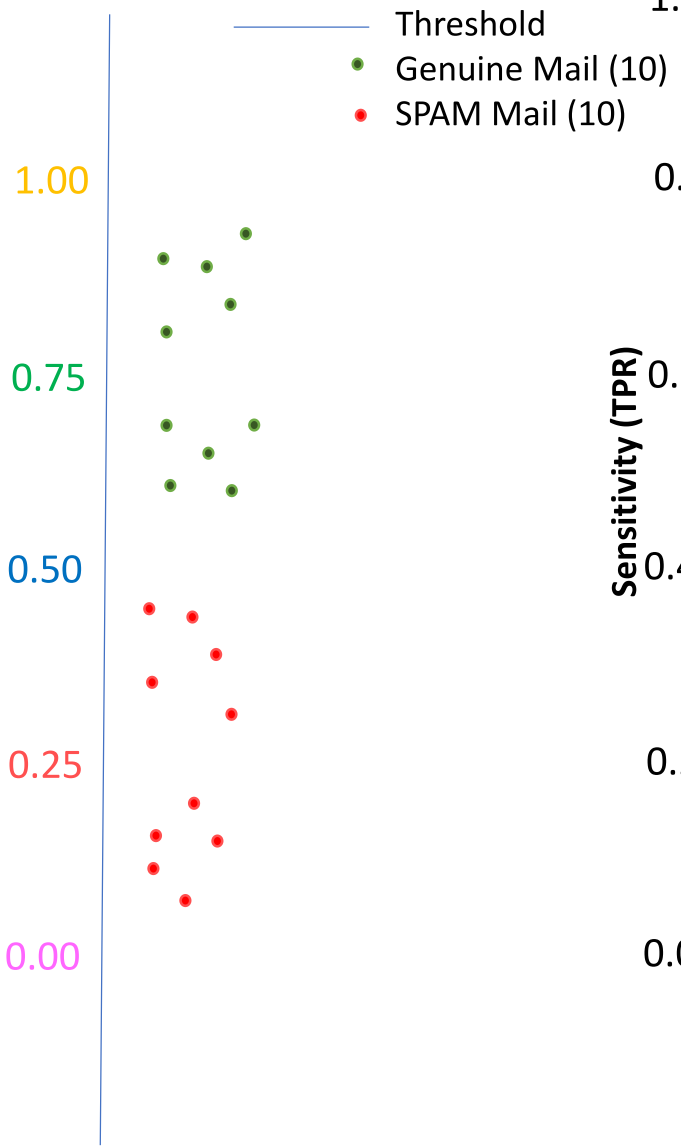
When threshold is near 0.25,
Sensitivity = TP/(TP+FN) = 10/10 = 1
Specificity = TN/(TN+FP) = 5/ 10 = 0.5
Hence (0.5,1)

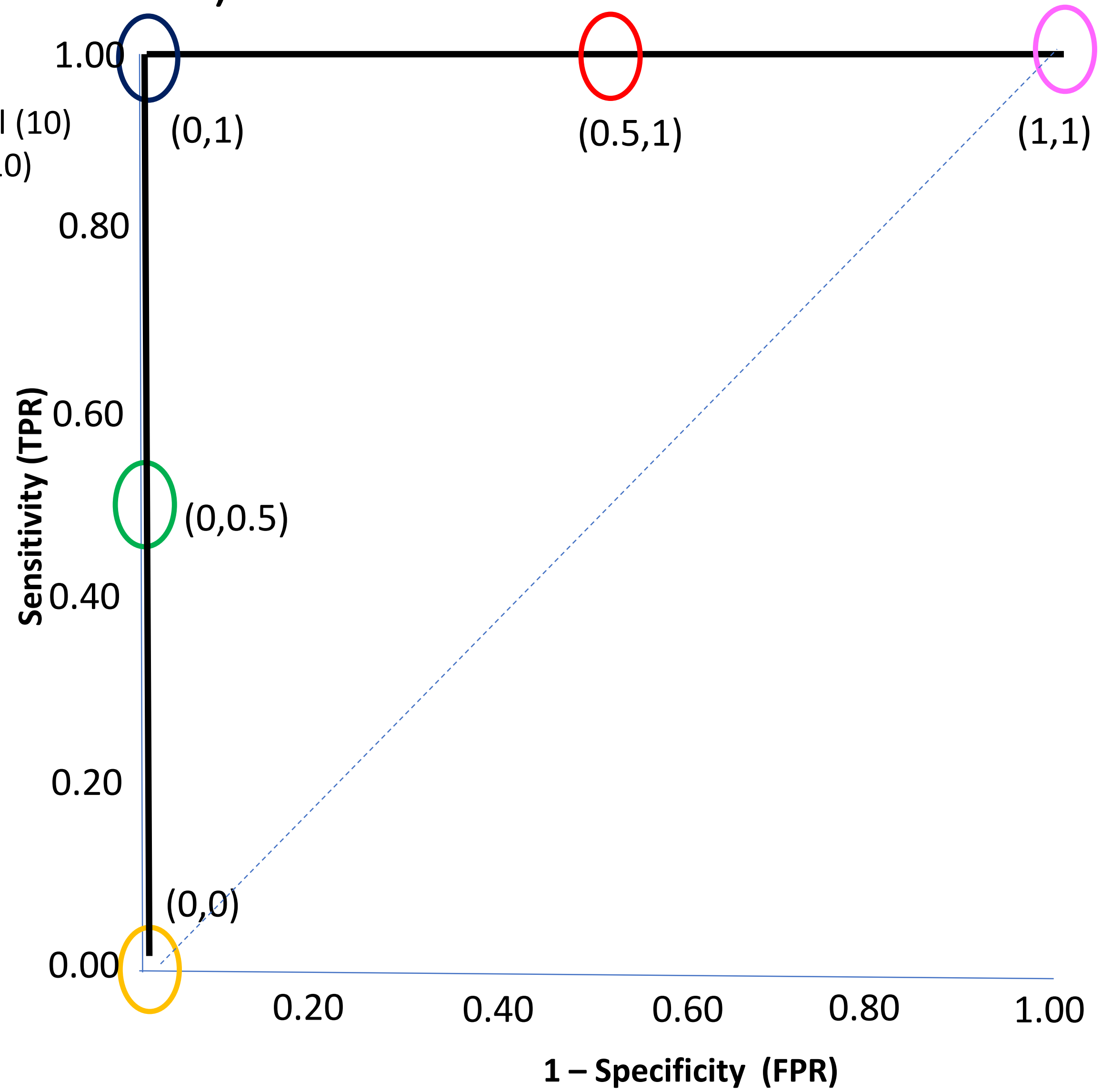
When threshold is near 0.0,
Sensitivity = TP/(TP+FN) = 10/10 = 1
Specificity = TN/(TN+FP)= 0/10 = 0
Hence (1,1)



ROC Curves (Receiver Operating Characteristic Curves) for a Perfect Classifier

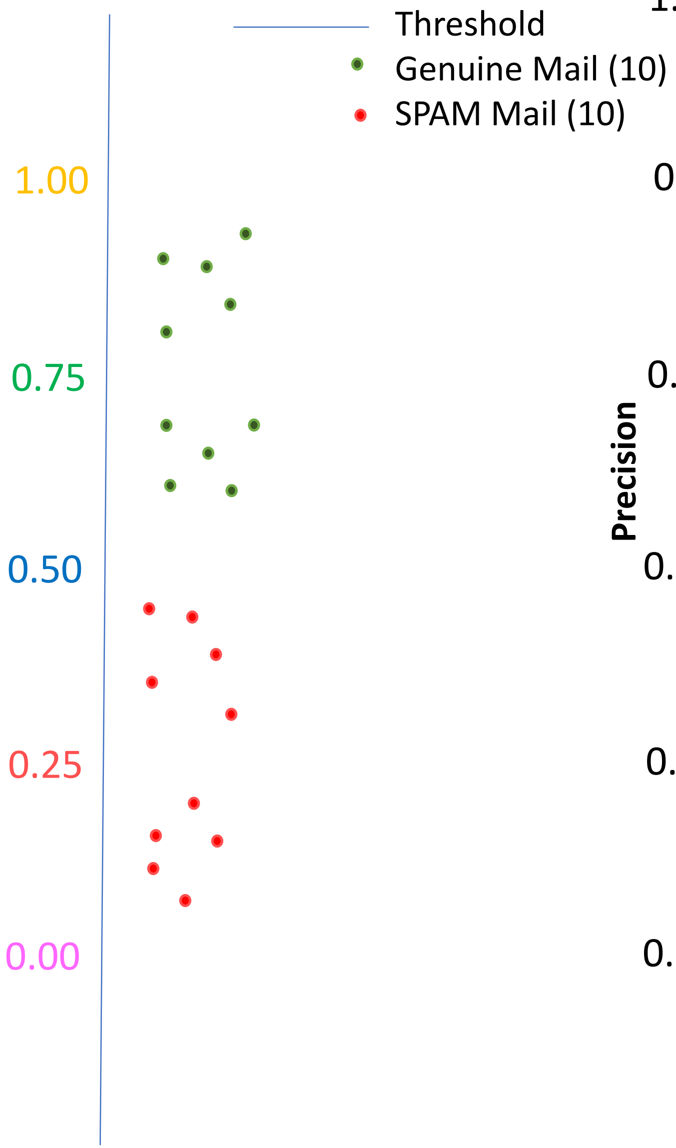
- ROC curve is an evaluation metric for binary classification problems.
- Plots the Sensitivity (**TPR)** against 1 – Specificity (**FPR)** for **varying** threshold values
- **Area Under the Curve (AUC)** is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve
- The goal in ROC space is to be in the ***upper-left-hand corner*** (0, 1), i.e. zero FP's (*FPR=0*) ; Sensitivity(Recall)=1, i.e. all the positives classified as positives
- The diagonal (the "*curve*" for a naïve classifier), on an average, when drawing randomly scores on the unit interval or if our predictions are all 0 or all 1.

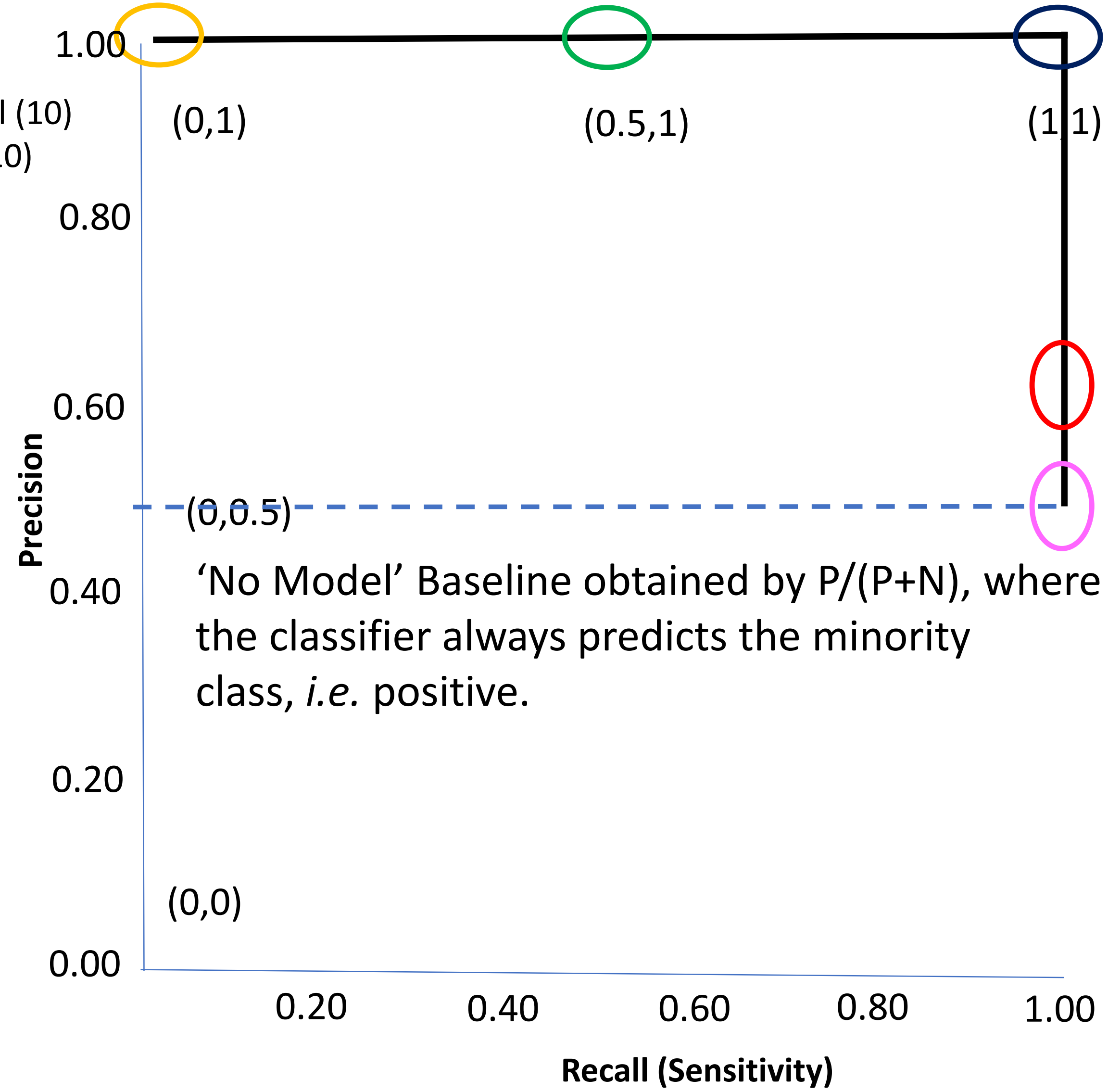




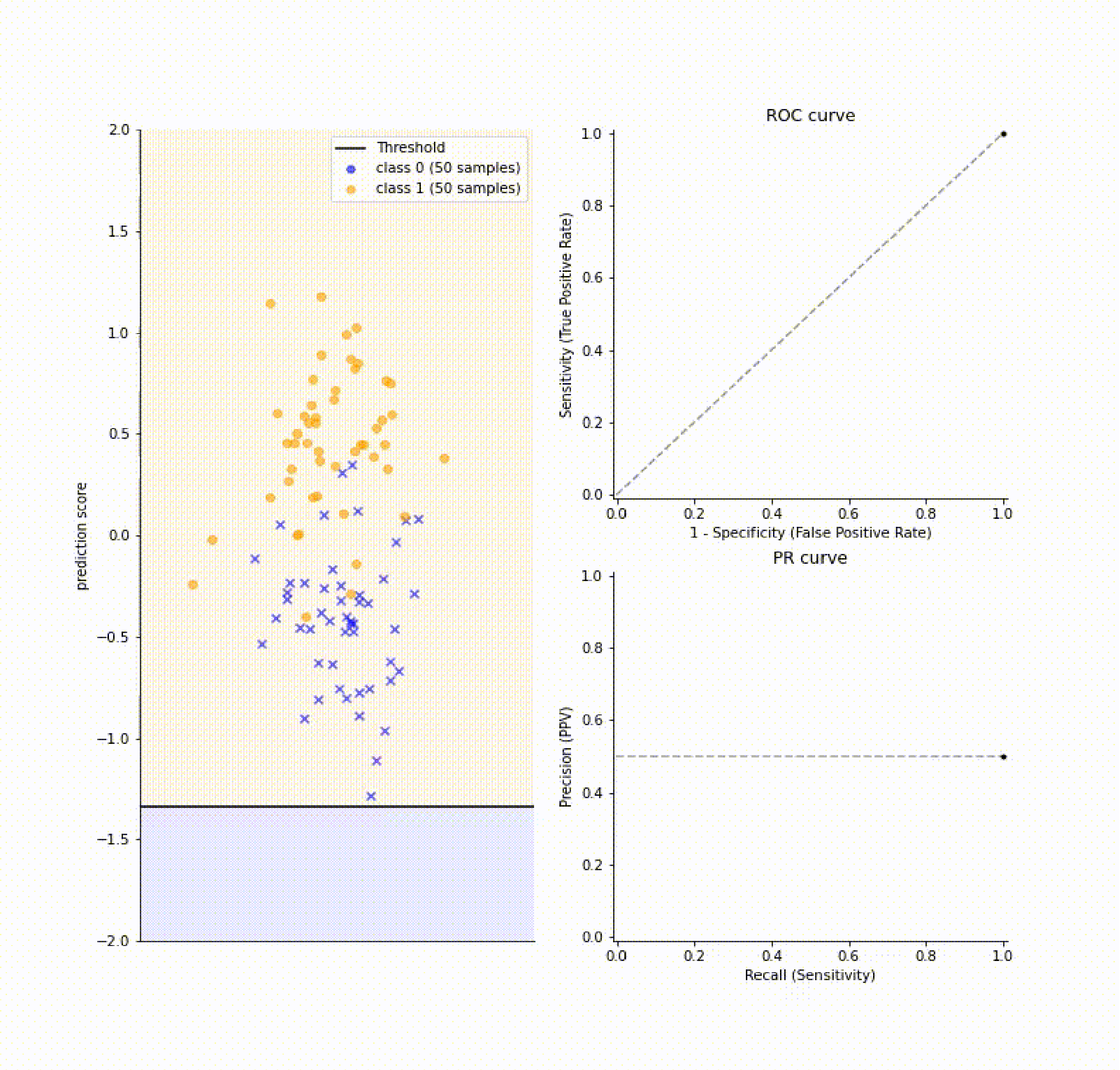
Precision Recall Curves for a Perfect Classifier

- Performance of the positive class (minority) when dealing with imbalanced classes.
- We are less interested in the skill of the model at predicting class 0 correctly, e.g. high true negatives. Key to the calculation of precision and recall is that the calculations do not make use of the true negatives. It is only concerned with the correct prediction of the minority class
- In the PR space, the goal is to be in the upper-right-hand corner — the top right corner (1, 1), all positives are classified as positive (*Recall=1*) and everything we are classifying as positive is true positive (*Precision=1*), ie zero False Positives.





ROC Curve for Imperfect classifier



- By design the ROC curve typically develops in a U-shape

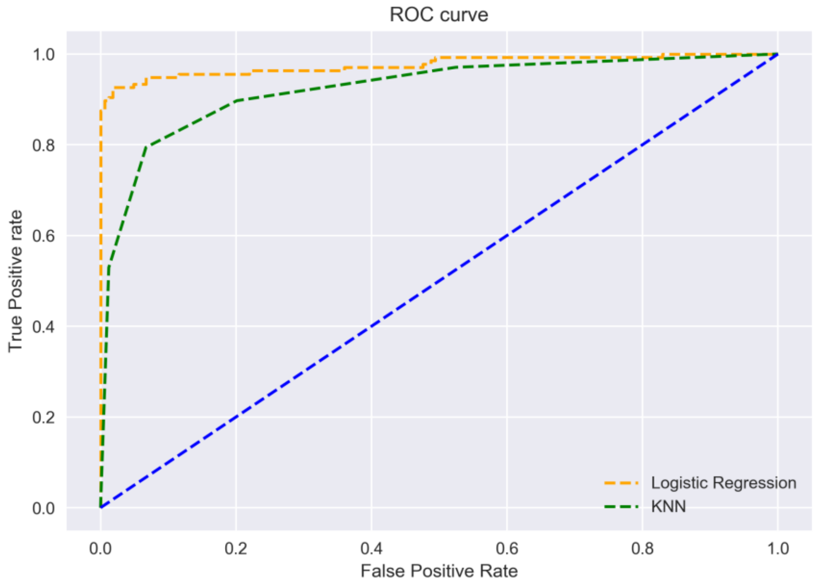
- **Step Left**: Turn a False Positive (blue cross) into a True Negative (blue dot). This behavior represents a correct decision and thus it reduces the False Positive Rate (*x-axis*) (mostly from -1.5 to -0.5)

- **Step downwards:** Turn a True Positive (orange dot) into a False Negative (orange cross). This behavior represents a wrong decision and thus it reduces the Recall (*y-axis*). (mostly from 0.5 to 1.5)

Image Source: https://towardsdatascience.com/on-roc-and-precision-recall-curves-c23e9b63820c

ROC curves for comparing models – Area Under the Curve

- The curves of different models can be compared directly in general or for different thresholds.
- As it can be challenging to compare two or more classifiers based on their curves, the Area Under the Curve (AUC) can be used as a summary of the model skill



- AUC for the Logistic Regression ROC curve is higher than that for the KNN ROC curve.
- Here Logistic regression performs better in classifying the positive class in this dataset

- PR AUC can be used for compare classification models on imbalanced datasets

Image Source: https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/

Summary

- ROC curve is an evaluation metric for binary classification problems.
- ROC plots the Sensitivity (**TPR)** against 1 – Specificity (**FPR)** for **varying** threshold values
- The goal in ROC space is to be in the ***upper-left-hand corner***
- ROC AUC used to compare across different classification models
- PR Curve is used to assess the performance of the model on positive class (minority) when dealing with imbalanced classes.
- PR AUC used to compare classification models on imbalanced datasets

References

- https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/
- https://towardsdatascience.com/on-roc-and-precision-recall-curves-c23e9b63820c
- https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/
- https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/