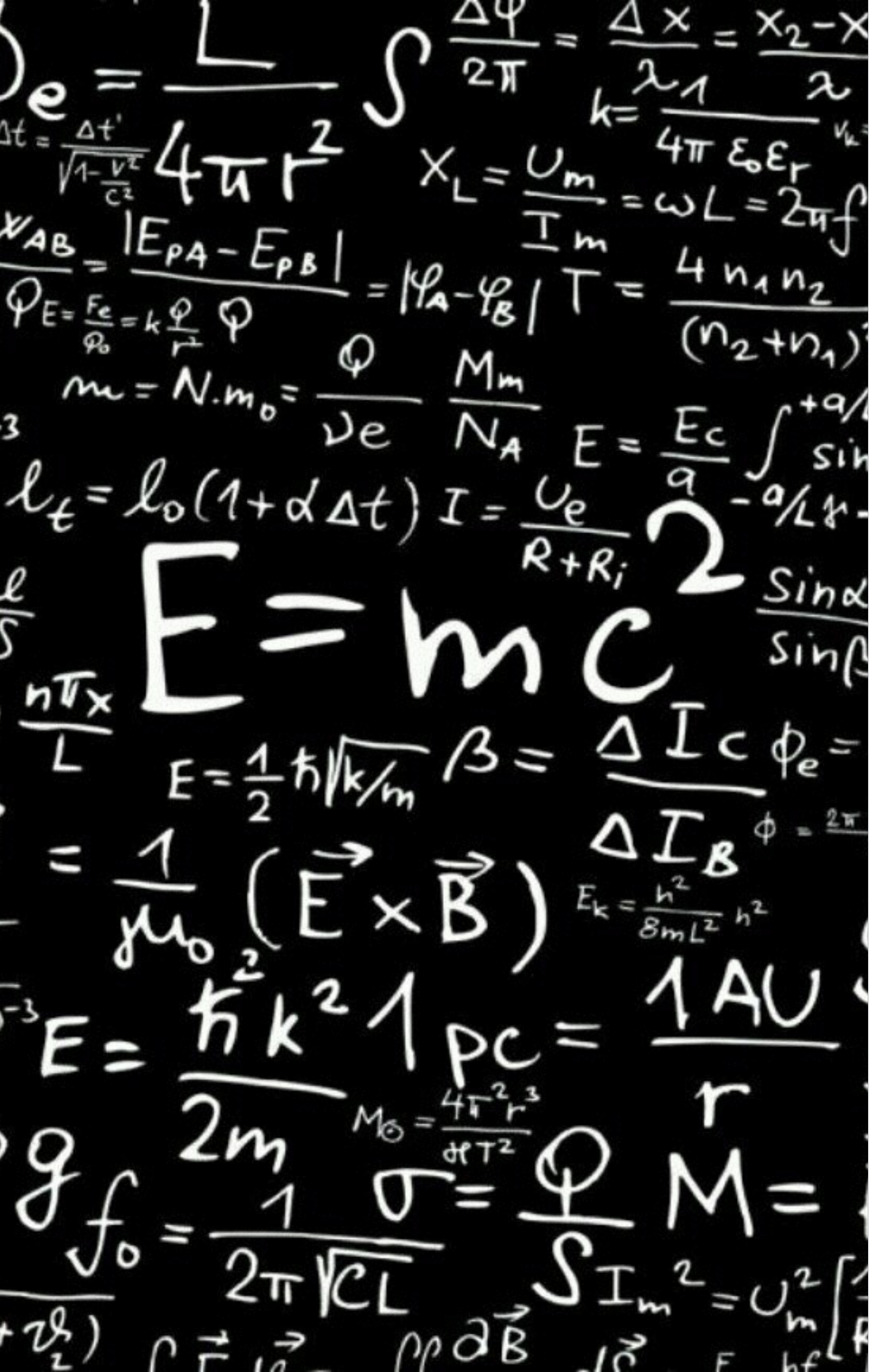


MODEL EVALUATIONS

Regression Evaluations

- Sum of Squared Error, Mean Squared Error, Root Mean Squared Error (SSE, MSE, RMSE)
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- R-Square
- Correlations

SSE, MSE, RMSE

$$SSE(h(x), y) = \sum_i (h(x_i) - y_i)^2$$

$$MSE(h(x), y) = \frac{1}{N} \sum_i (h(x_i) - y_i)^2$$

$$RMSE(h(x), y) = \frac{1}{N} \sum_i \sqrt{(h(x_i) - y_i)^2}$$

MAE, MAPE

- Mean absolute error is very similar to root mean square error. Can be viewed as L1 norm of loss (while squared error is L2 norm)

$$MAE(h(x), y) = \frac{1}{N} \sum_i |h(x_i) - y_i|$$

- Mean absolute percentage error give you percent errors which are easier to interpret, but sensitive to small targets.

$$MAPE(h(x), y) = \frac{1}{N} \sum_i |\frac{h(x_i) - y_i}{y_i}|$$

R-Squared

- R-squared (Coefficient of Determination) : how close the data are to the fitted regression line.
- R-squared = the percentage of the response variable variation that is explained by a linear model.

$$\text{R-squared} = 1 - \frac{\text{Unexplained Variance}}{\text{Total Variance}}$$

- R-squared = 0 (model explains none of the variability of data).
- R-squared = 1 (model explains all of the variability of data).

R-Squared v.s. Correlations

- R-Squared measures how much variance of y is explained by predictions.

$$R^2(h(x), y) = 1 - \frac{\sum_i (y_i - h(x_i))^2}{\sum_i (y_i - \bar{y})^2}$$

← Unexplained Variance (errors)

← Total Variance

- Correlations measures how much variance of y and prediction are varying together.

$$\rho(\hat{y}, y) = \frac{E[(\hat{y} - \mu_{\hat{y}})(y - \mu_y)]}{\sigma_{\hat{y}} \sigma_y}$$

Regression Evaluations

- Sum of Squared Error, Mean Squared Error, Root Mean Squared Error (SSE, MSE, RMSE) - easy for optimization algorithm due to the differentiable form.
- Mean Absolute Error (MAE) - very similar to RMSE. Might work better for discrete y.
- Mean Absolute Percentage Error (MAPE) - easy to interpret for most people.
- R-Square - give you a sense of how much your model has done to explain y, how much room to improve.
- Correlations - easy to beat, since it doesn’t require y and predictions to have the same scale.

Classification Evaluations

- Log Loss
- Accuracy
- Precision, Recall, F1-Score
- Precision recall curve
- Average precision
- ROC

Classification Performance

- Log Loss
- Accuracy
- Precision
- Recall
- F1-Measure

Predicted class

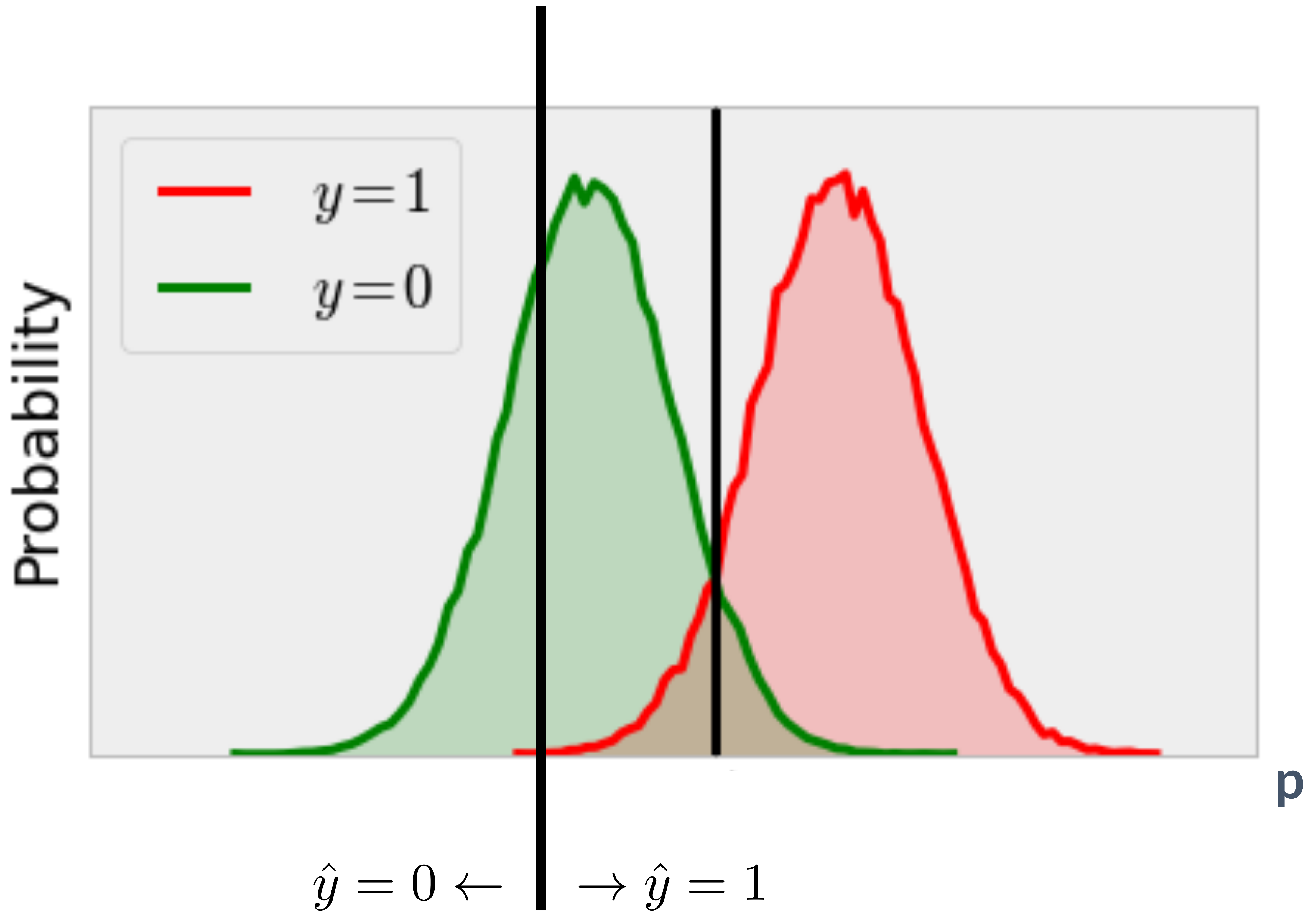
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

$$PRE = \frac{TP}{TP + FP}$$

$$REC = TPR = \frac{TP}{P} = \frac{TP}{FN + TP}$$

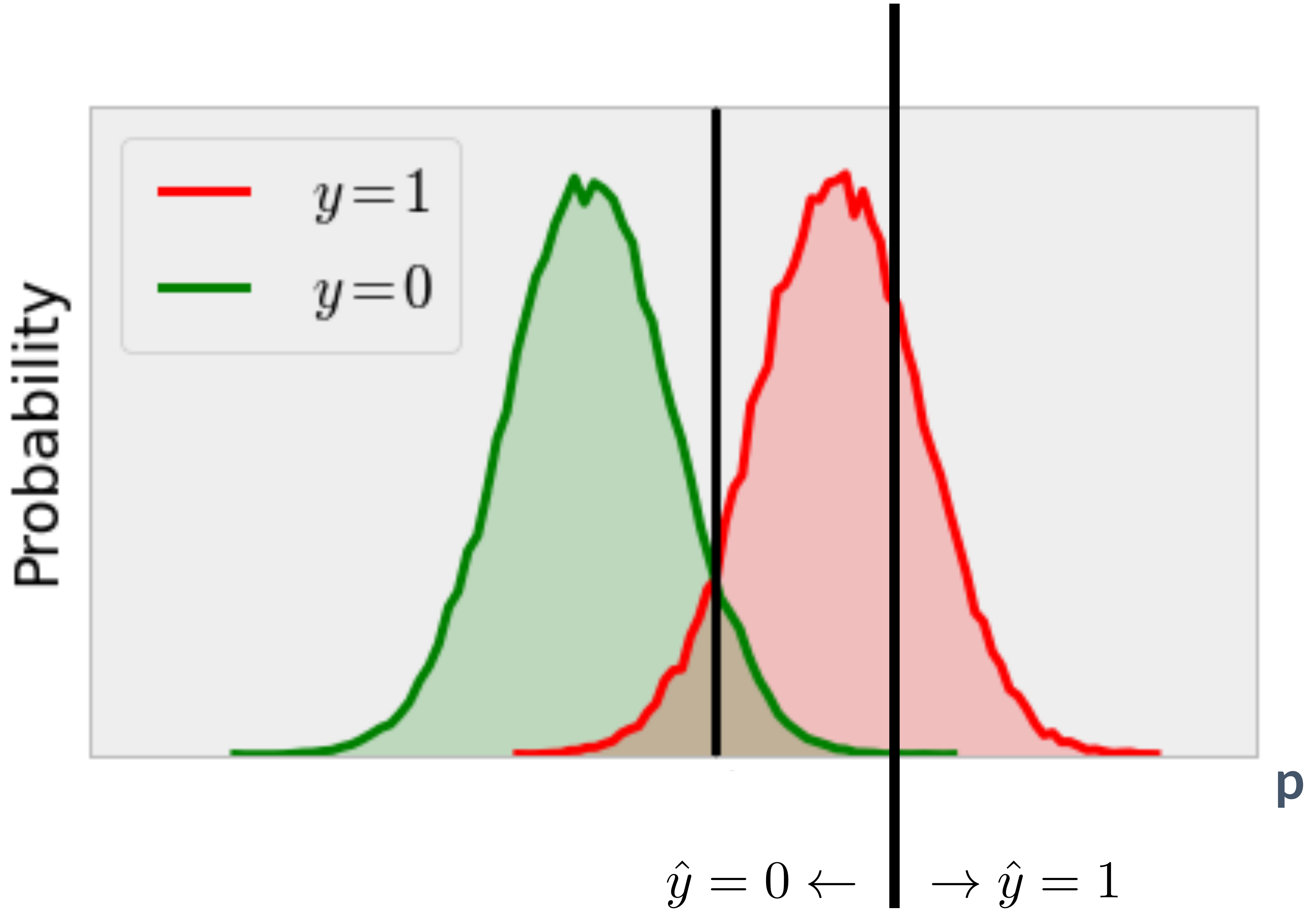
$$F_1 = 2 \cdot \frac{PRE \cdot REC}{PRE + REC}$$

Precision Recall Tradeoff



- Precision recall obviously depends on your probability threshold (p=0.5 is default, but you can use other numbers).
- The lower threshold you use, the higher recall (you are going to retrieve more positive samples by relaxing your criteria)
- However, you lost precision because you are too relaxed, letting more people become positive samples.

Precision Recall Tradeoff

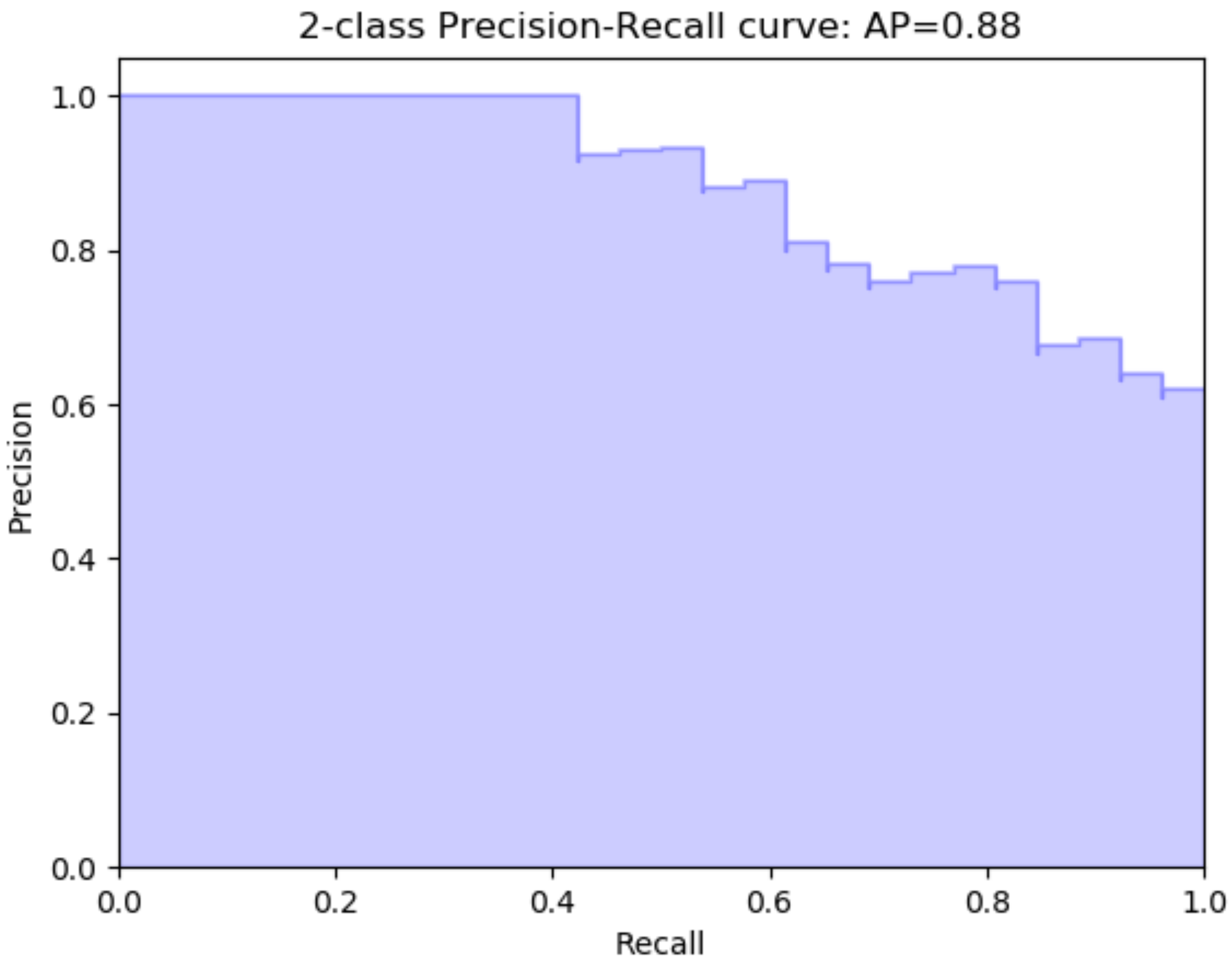


- As threshold gets higher, meaning you are extremely selective, it's likely you are going to get true positives.
- However, you will miss out a lot of other positive samples, because not many sample will get pass such a strict criteria.

Precision Recall Curve

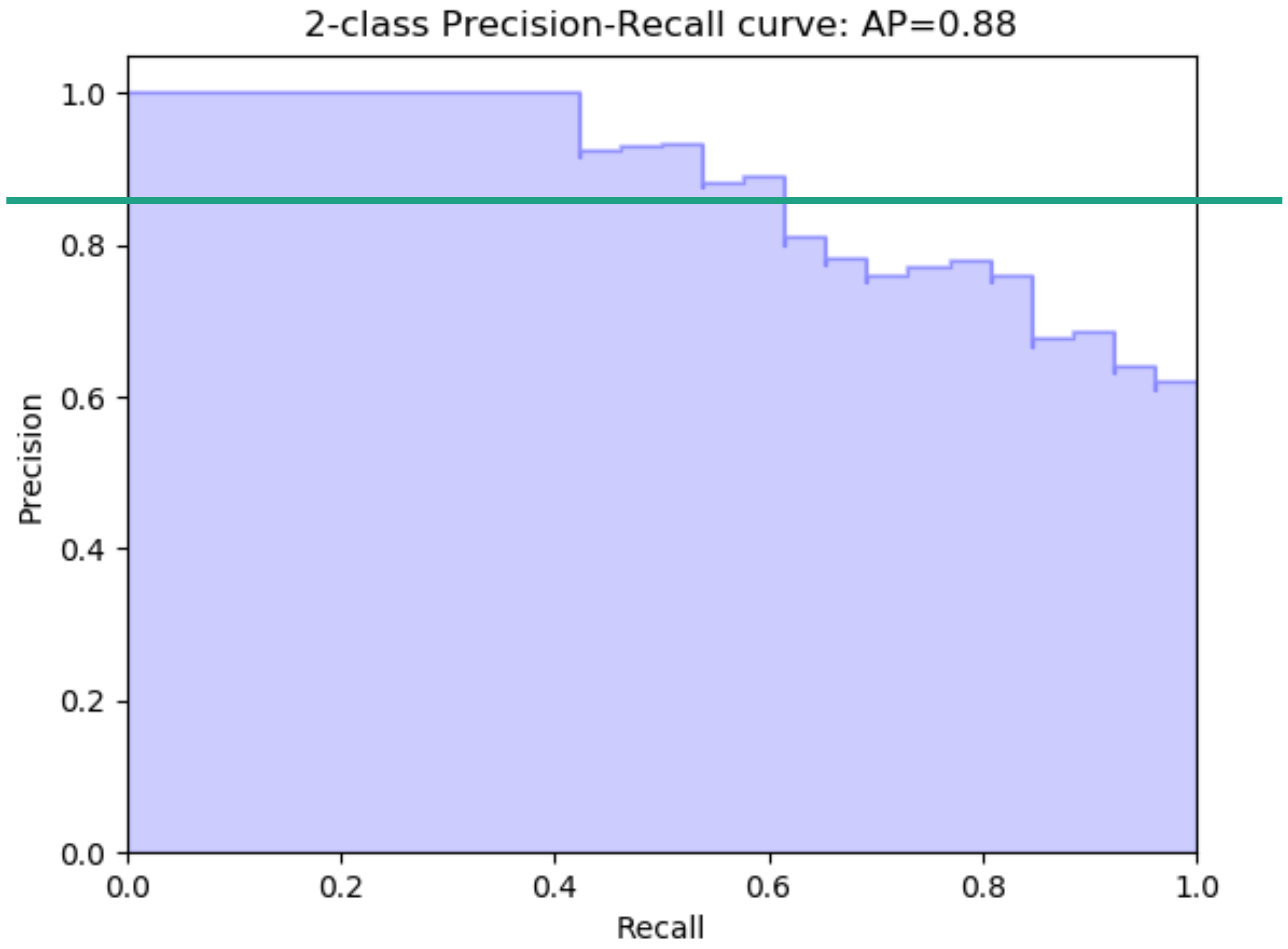
- The precision-recall curve shows the tradeoff between precision and recall for different threshold.
- A high area under the curve represents both high recall and high precision (low false positives and negatives)
- High scores for both show that the classifier is returning accurate results (high precision),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

Precision Recall Curve



- Precision recall curve plots precision and recall at various thresholds.
- High precision / low recall points correspond to high thresholds.
- Low precision / high recall points correspond to low thresholds.
- Large blue area means your classifier is awesome.

Average Precision

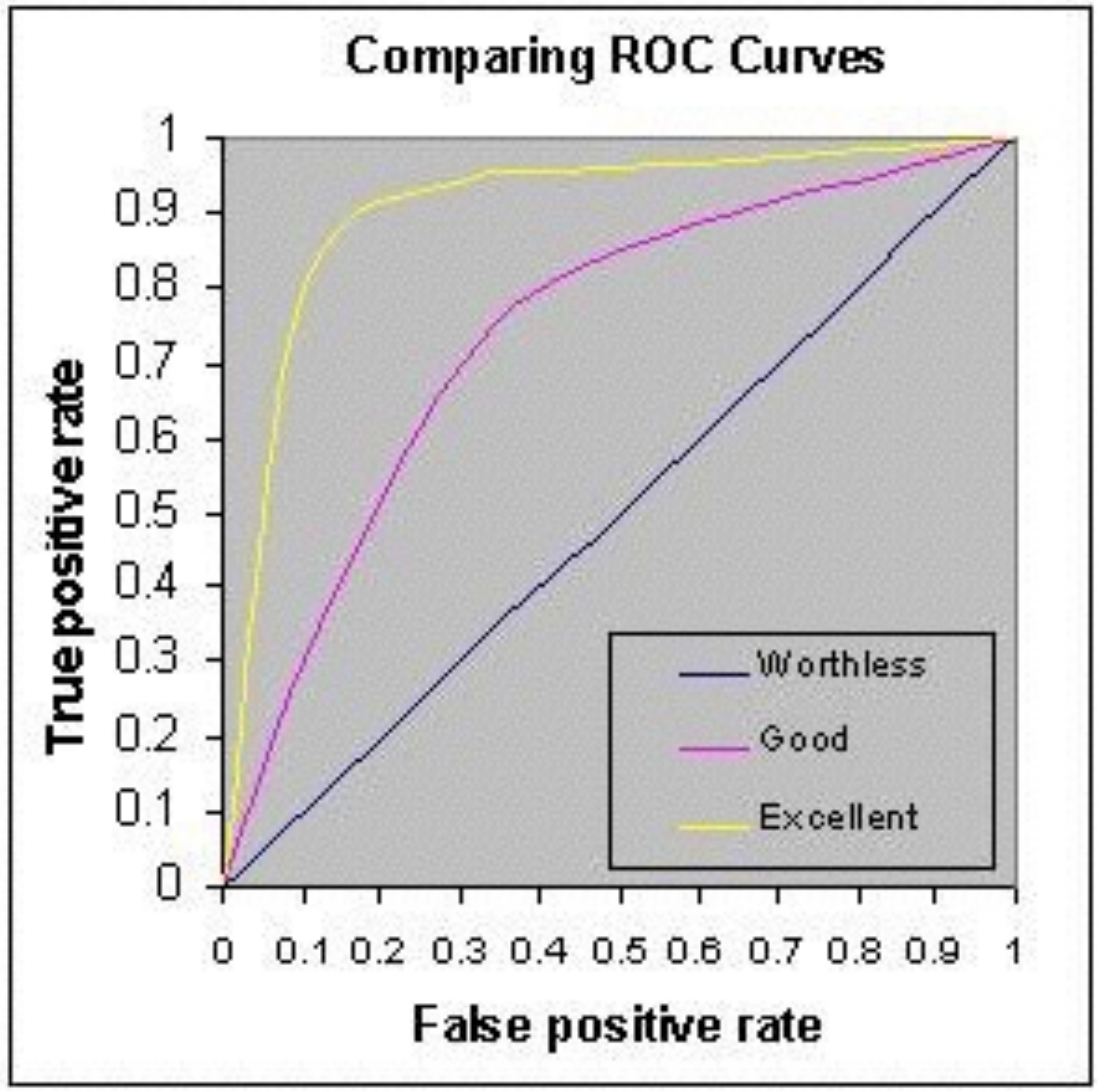


- Average Precision summarizes precision recall tradeoff
- AP = weighted mean of precisions achieved at each threshold
- weight = the increase in recall from the previous threshold

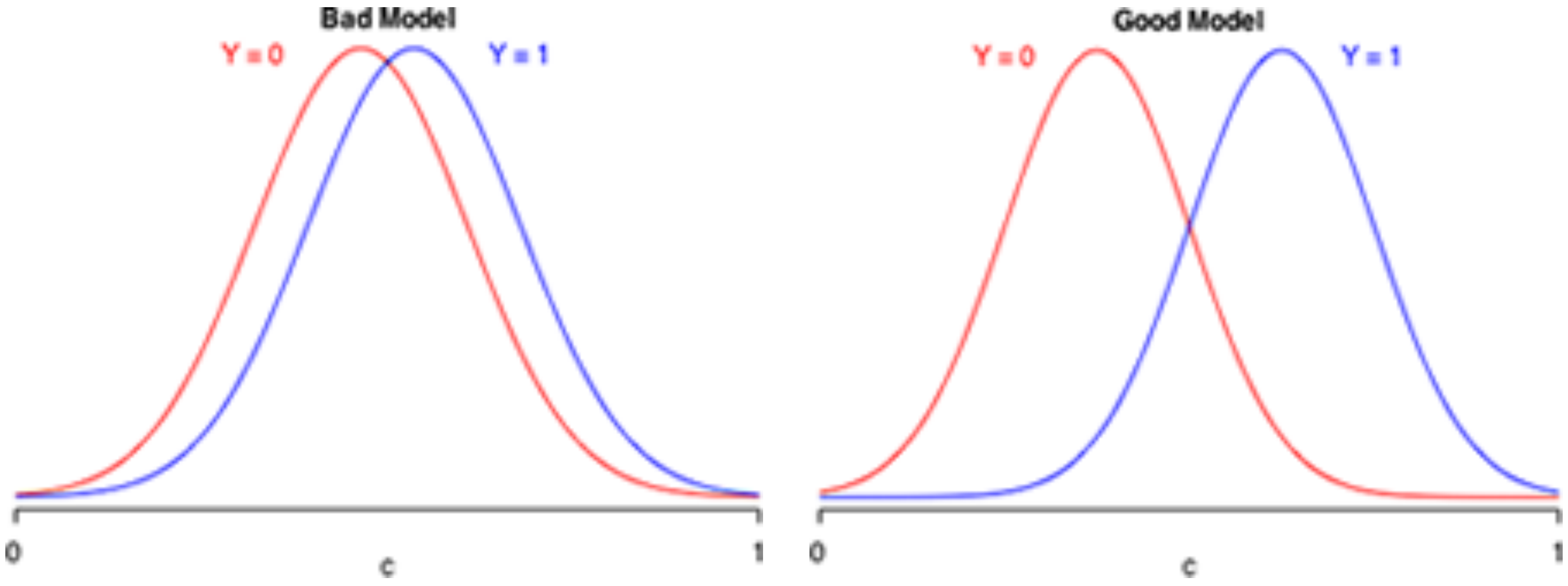
$$AP = \sum_{n} (R_n - R_{n-1}) P_n$$

ROC - Receiver Operating Characteristic

- Similar to precision-recall curve ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at various thresholds.



- It measures how good h(x) is as a differentiator of true classes, y.



Classification Evaluations

- Log Loss - differentiable, easy for optimization algorithm
- Accuracy - easy to interpret
- Precision, Recall, F1-Score - use this instead of accuracy for imbalanced classification.
- Precision recall curve - give you a more clear picture of imbalanced classification problem
- Average precision - summarizes precision recall curve
- ROC - similar to precision-recall curve, but often more sensitive (easier to beat)