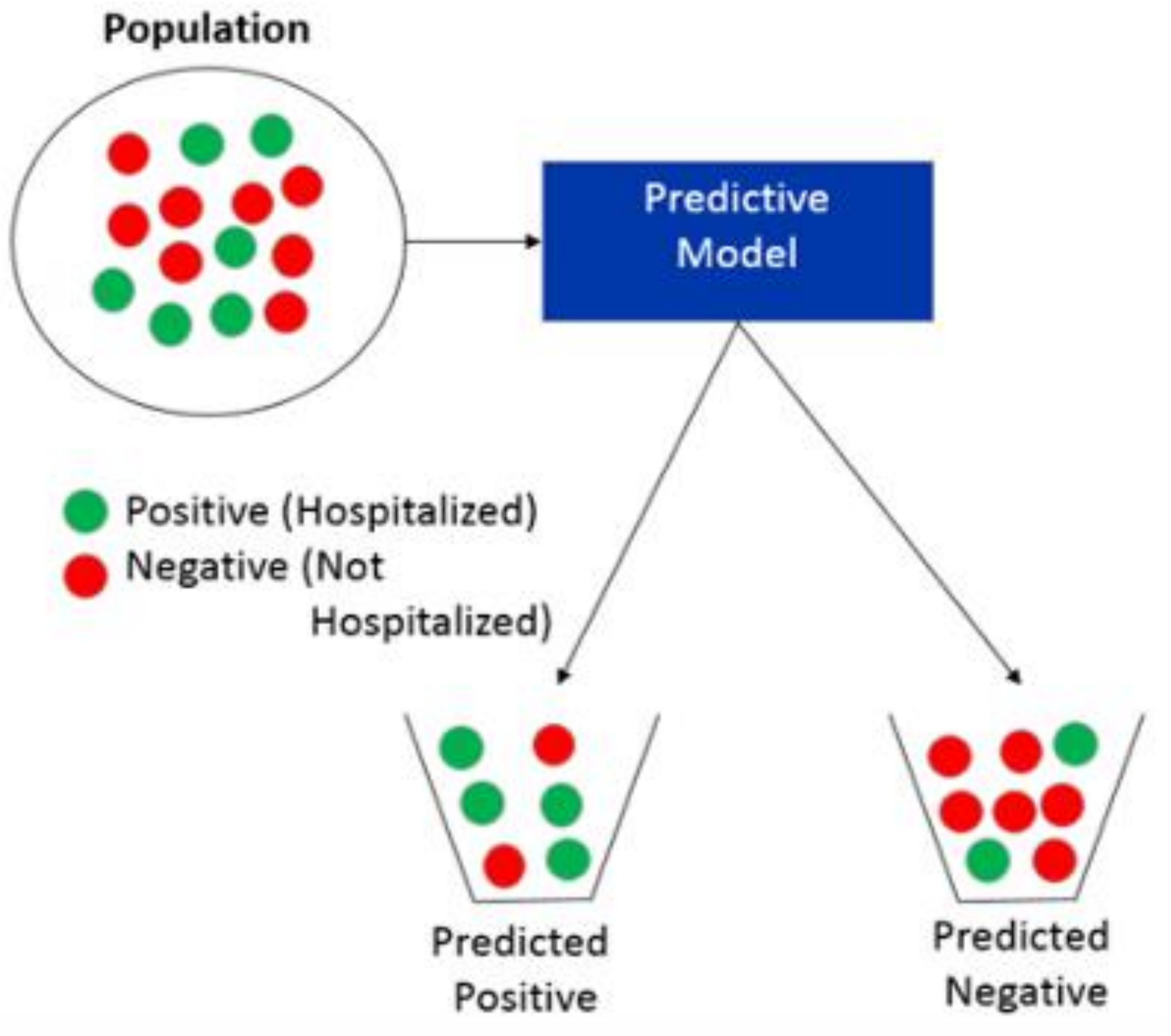
Classification Metrics


Population
Predictive Model
Positive (Hospitalized)
Negative (Not Hospitalized)
Predicted Positive
Predicted Negative

Confusion Matrix

Given an actual label and a predicted label, the first thing we can do is divide our samples in 4 buckets:

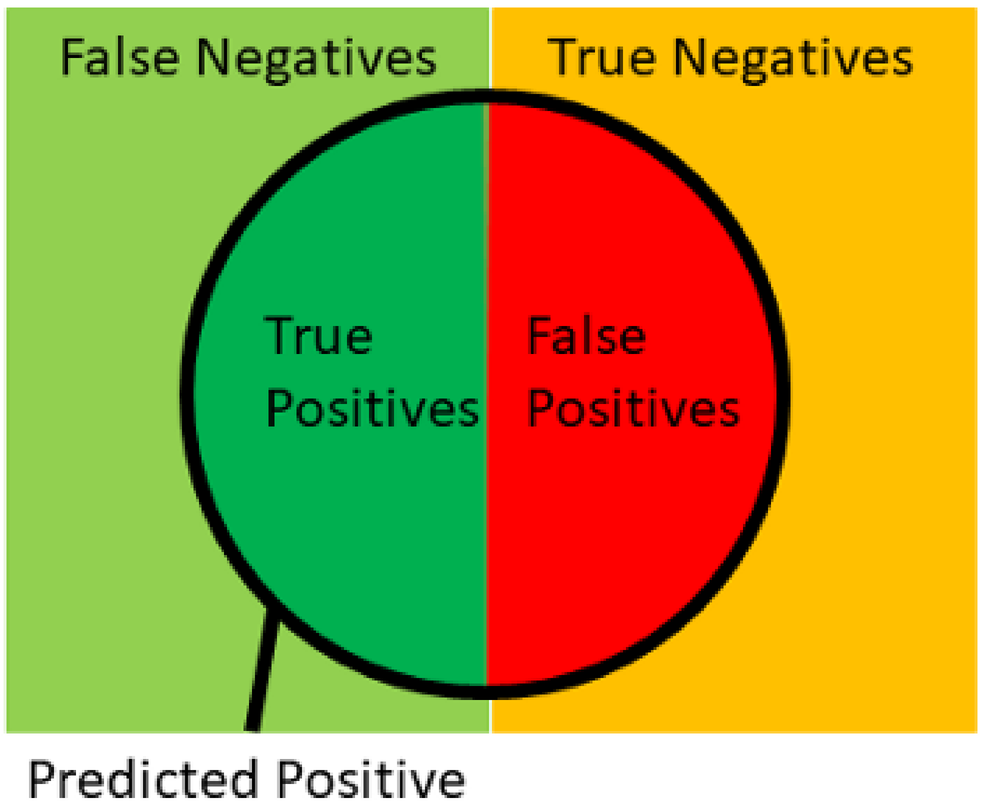
True positive — actual = 1, predicted = 1

False positive — actual = 0, predicted = 1

False negative — actual = 1, predicted = 0

True negative — actual = 0, predicted = 0

Confusion Matrix

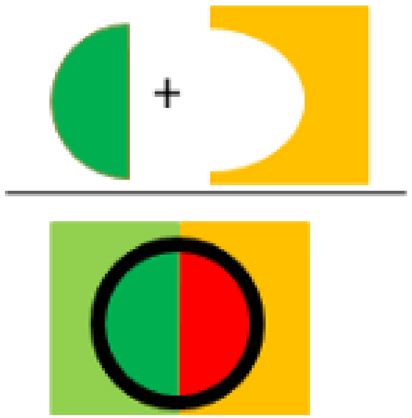


Confusion Matrix		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Accuracy Score

- The most common metric for classification is accuracy, which is the fraction of samples predicted correctly as shown below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} =$$

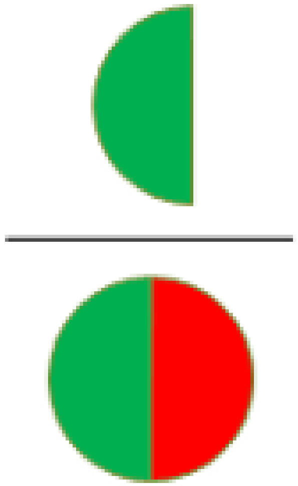


Fraction predicted correctly

Precision Score

- Precision is the fraction of predicted positives events that are actually positive as shown below:

$$\text{Precision} = \frac{TP}{TP + FP} =$$

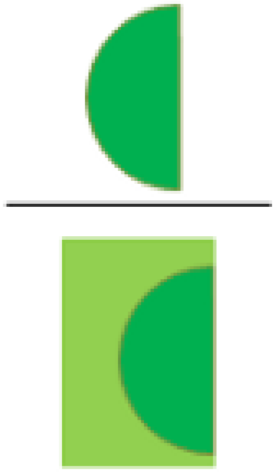


Fraction of predicted positives that are actually positive

Recall Score

- Recall (also known as sensitivity) is the fraction of positives events that you predicted correctly as shown below:

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} =$$



Fraction of positives predicted correctly

Precision vs Recall

- For example, a cancer hospital has 200 patients where 100 have cancer
- A Machine Learning model is retrieving 80 patients and saying that those 80 have cancer and rest of 120 don’t have cancer
- Imagine that 40 people out of the 80 patients retrieved by the model have cancer
- Precision is 40/80 = 0.5, Recall is 40/100 = 0.4
- Now which one is important?

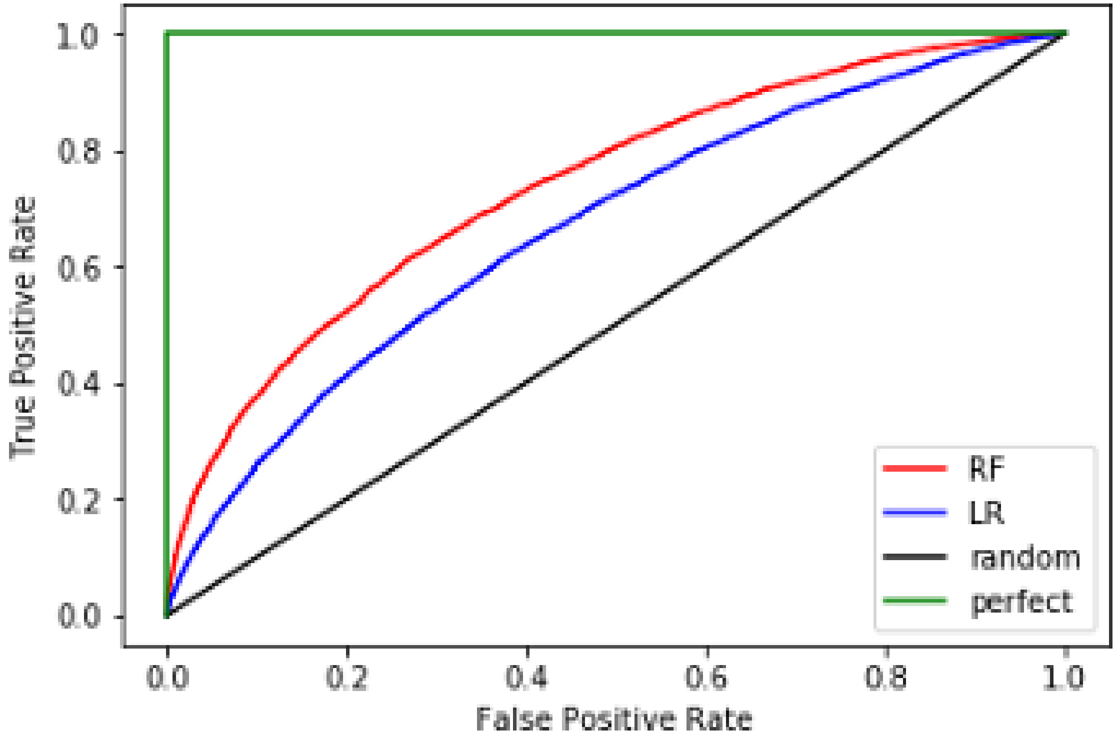
F1 Score

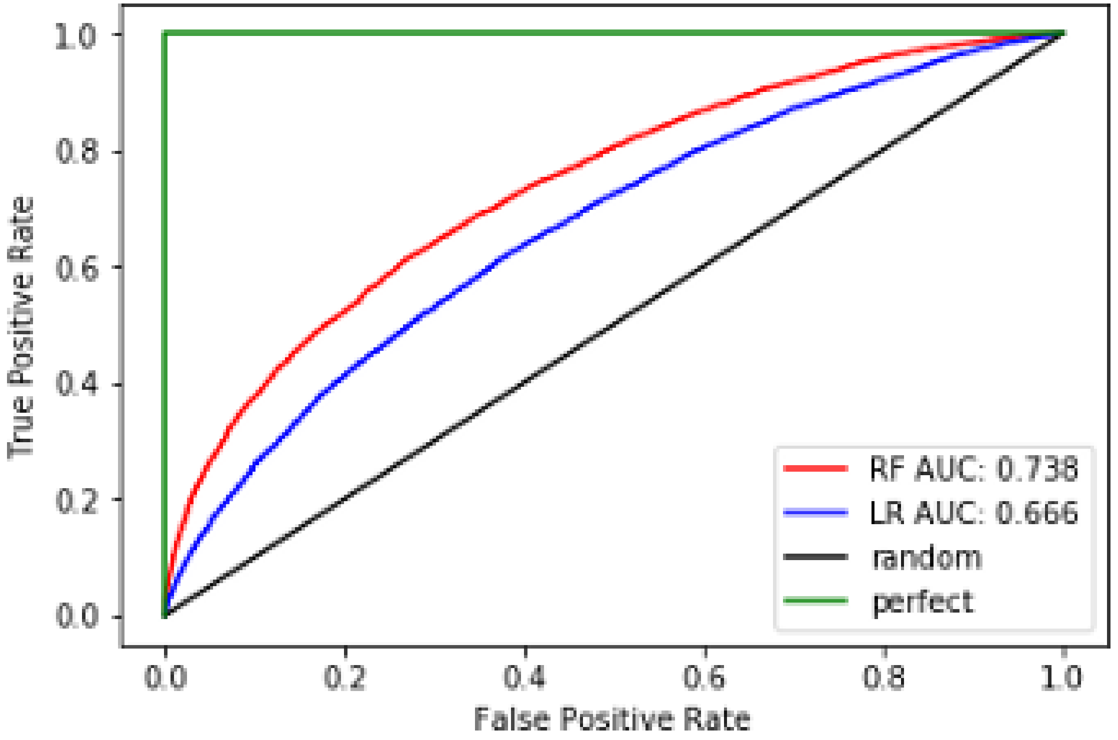
- The f1 score is the harmonic mean of recall and precision, with a higher score as a better model. The f1 score is calculated using the following formula:

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * (precision * recall)}{precision + recall}$$

ROC Curve and ROC AUC Score

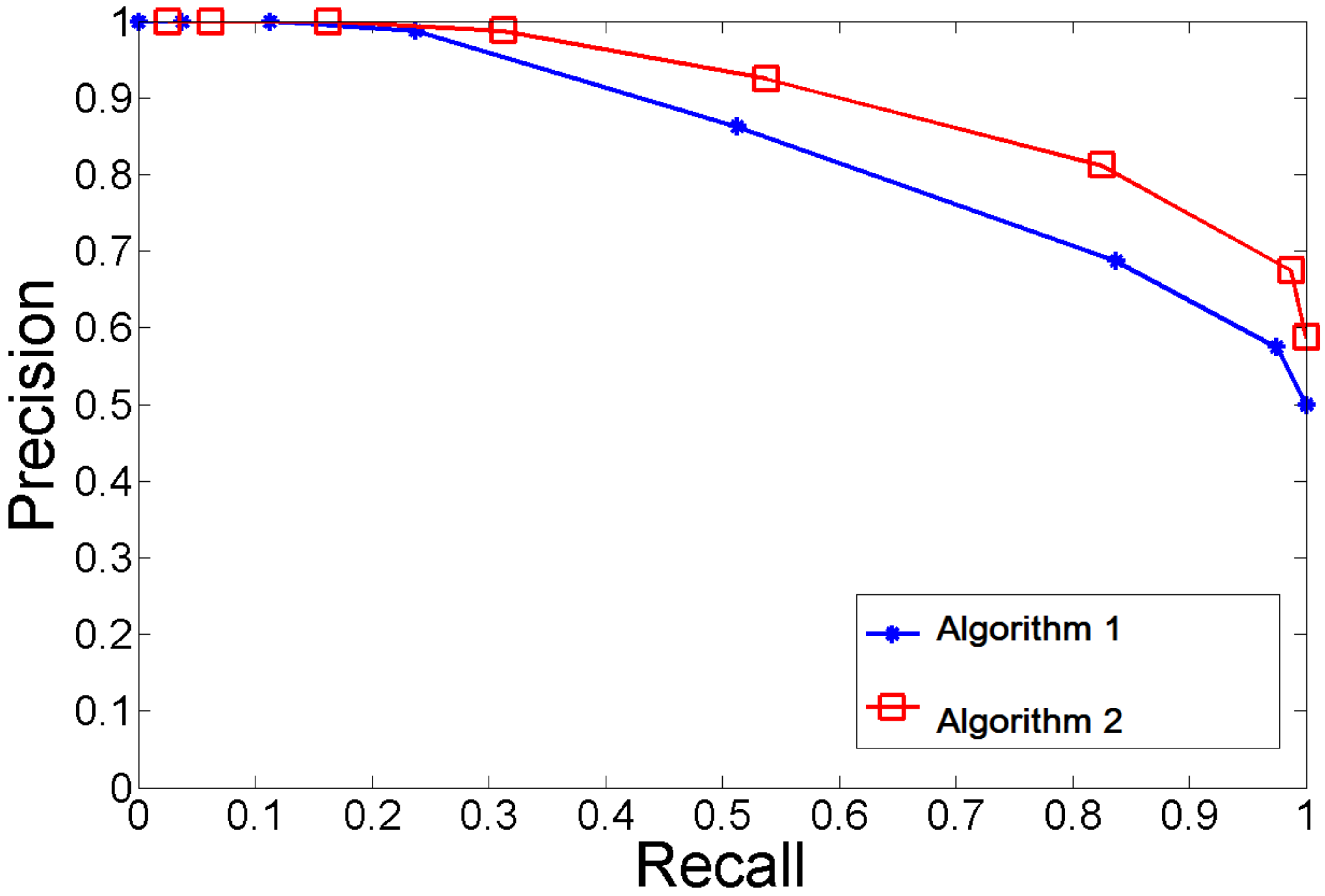
- ROC curves are VERY help with understanding the balance between true-positive rate and false positive rates. Calculated using 3 lists
- thresholds = all unique prediction probabilities in descending order
- fpr = the false positive rate (FP / (FP + TN)) for each threshold
- tpr = the true positive rate (TP / (TP + FN)) for each threshold



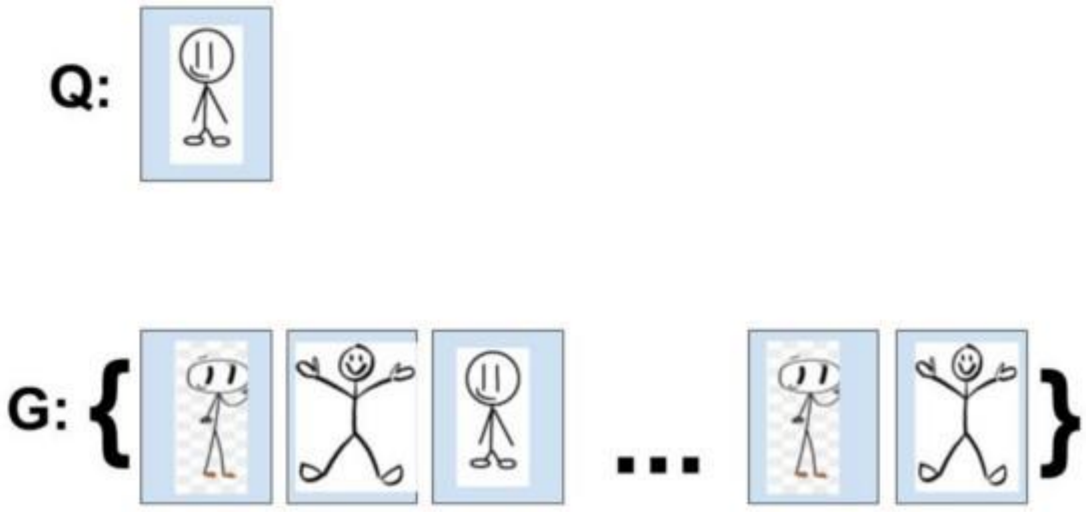


ROC Curve and ROC AUC Score

Precision Recall Curve

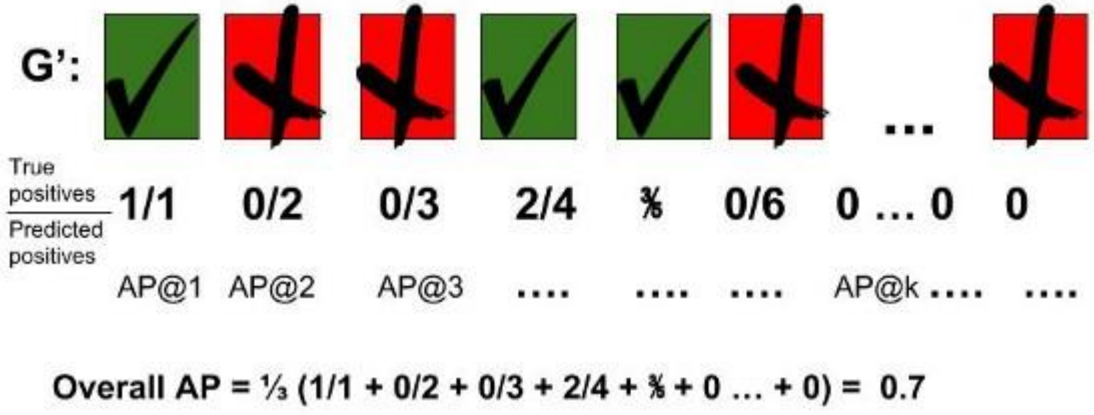


Average Precision

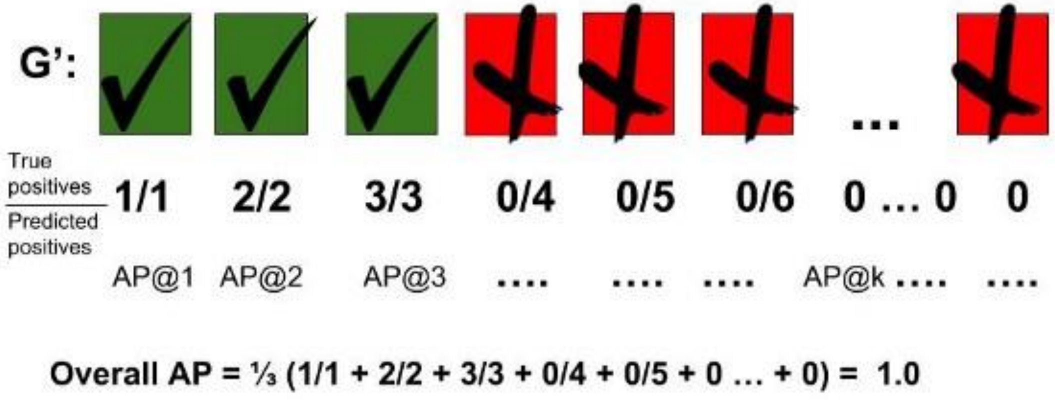


$$\text{AP@k} = \frac{1}{\text{GTP}} \sum_{i=1}^{k} \frac{\text{TP seen}}{i}$$

AP@k formula for information retrieval tasks



Calculation of AP for a given query, Q, with a GT=3



Calculation of a perfect AP for a given query, Q, with a GTP=3

Mean Average Precision (mAP)

- For each query, Q, we can calculate a corresponding AP. A user can have as much queries as he/she likes against any labeled database. The mAP is simply the mean of all the queries that the use made.

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i$$

mAP formula for information retrieval

Intersection over Union (IoU)

- Useful in object detection and image segmentation problems