Topic: Performance Metrics

Performance Metrics

- Performance metrics are a part of every machine learning pipeline.
- They are used to monitor and measure the performance of a model (during training and testing).
- They tell you if you’re making progress, and put a number on it.
- Helpful to decide the best model to meet the target performance.
- All machine learning models, whether it’s linear regression, or a SOTA technique need a metric to judge performance.

Types of Performance Metrics

In order to evaluate Classification models, following metrics are used:

- Accuracy
- Confusion Matrix (not a metric but fundamental to others)
- Precision
- Recall
- F1-score
- Intersection Over Union (IoU)
- AU-ROC

Accuracy

- Classification accuracy is perhaps the simplest metric to use and implement and is defined as the number of correct predictions divided by the total number of predictions, multiplied by 100.

- We can implement this by comparing ground truth and predicted values In a loop or simply utilizing the scikit-learn module.

Accuracy = (no. of correct predictions ÷ total no. of predictions)*100

Confusion Matrix

Confusion Matrix is a tabular visualization of the actual values labels versus predicted values. Each row of the confusion matrix represents the instances in a predicted class and each column represents the instances in an actual class.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

There are 4 important terms :

- True Positives (TP) : The cases in which we predicted YES and the actual output was also YES.

- True Negatives (TN) : The cases in which we predicted NO and the actual output was NO.

- False Positives (FP): The cases in which we predicted YES and the actual output was NO.

- False Negatives (FN): The cases in which we predicted NO and the actual output was YES.

Precision

Precision is the ratio of true positives and total positives predicted:

A precision score towards 1 will signify that your model didn't miss any true positives, and is able to classify well between correct and incorrect labeling.

A low precision score (<0.5) means your classifier has a high number of false positives which can be an outcome of imbalanced class or untuned model hyper parameters

precision = TP÷(TP+FP)

Recall

- A Recall is essentially the ratio of true positives to all the positives in actual values.

- Recall towards 1 will signify that your model didn't miss any true positives, and is able to classify well between correctly and incorrectly labeling of objects.

- A low recall score (<0.5) means your classifier has a high number of false negatives which can be an outcome of imbalanced class or untuned model hyper parameters.

- Recall = TP÷(TP+FN)

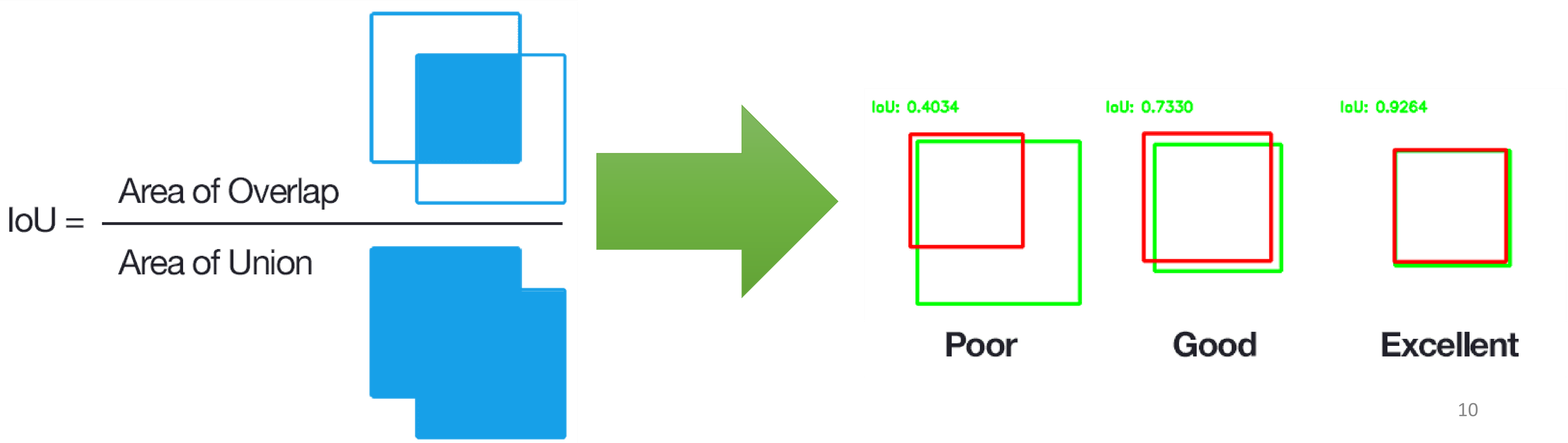
F1-Score

- The F1-score metric uses a combination of precision and recall. In fact, the F1 score is the harmonic mean of the two.
- The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.
- It ranges from 0 to 1, where 1 indicates perfect precision and recall, and 0 means neither precision nor recall.

F1-score = 2*(Recall * Precision) ÷(Precision + Recall)

Intersection Over Union (IoU)

IoU measures the overlap between the predicted bounding box and the actual bounding box. It is calculated as the Ratio of the Intersection of the two bounding boxes to the Union of the two bounding boxes.



Area Under Receiver Operating Characteristics Curve (AUROC)

- Better known as AUC-ROC score/curves. It makes use of true positive rates(TPR) and false positive rates(FPR).
- AUC - ROC curve is a performance measurement for the classification problems at various threshold settings.
- ROC is a probability curve and AUC represents the degree or measure of separability.
- It tells how much the model is capable of distinguishing between classes.
  - Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1.
  - Higher the AUC, the better the model is at distinguishing for example patients with the disease and no disease.
- The ROC curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on the x-axis.

AUROC

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

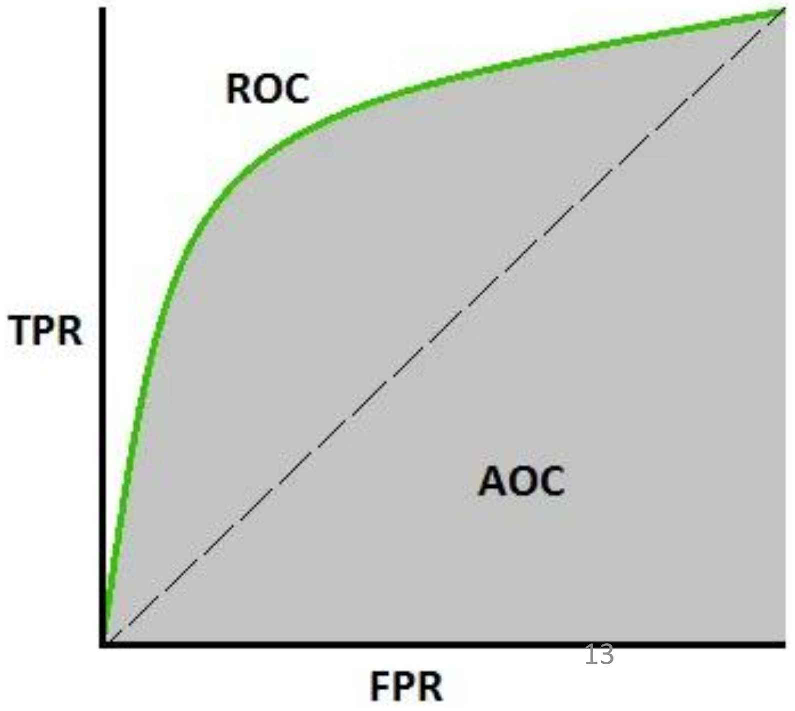
$$\text{FPR} = 1 - \text{Specificity}$$

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

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

AUROC

To combine the FPR and the TPR into a single metric, we first compute the two former metrics with many different thresholds for the logistic regression, then plot them on a single graph. The resulting curve is called the ROC curve, and the metric we consider is the area under this curve, which we call AUROC.



Thank You