Performance Metrics for Classification, Detection Problems

Confusion Matrix

- Confusion matrix is one of the basic table layout that helps **to visualizes the performance** of a classification and detection related machine learning algorithms.

		True Class Label	
Predicted Class Label		Positive	Negative
	Positive	***TP***	***FP***
	Negative	***FN***	***TN***

Testing

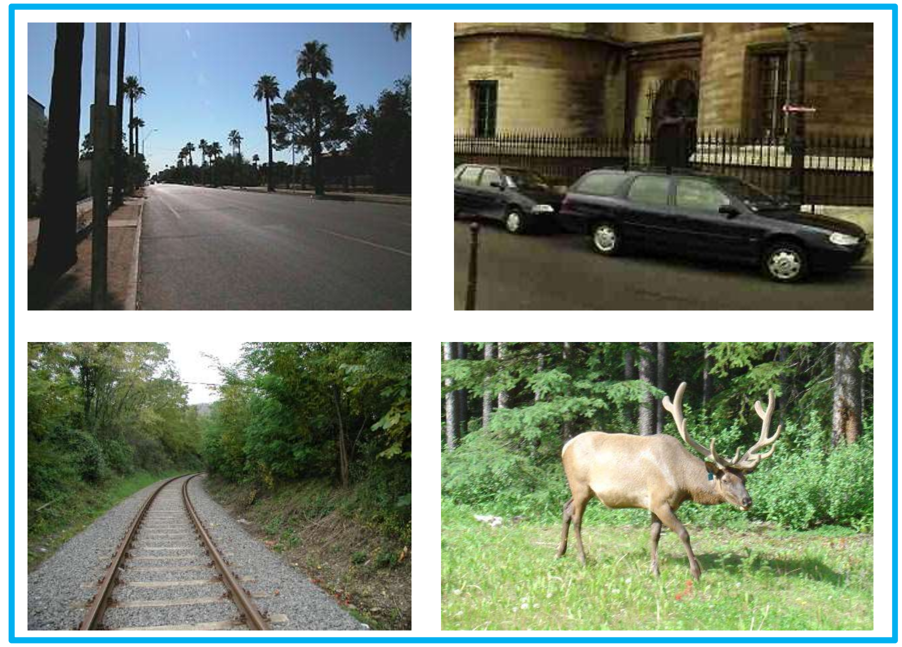
Test : Which objects are blue?

Objects: ,

Object	Test Result	Think	What is this?
	Blue	Correctly identified	True Positive
	Blue	Incorrectly identified	False Positive
	Not Blue	Incorrectly rejected	False Negative
	Not Blue	Correctly rejected	True Negative

Human detection framework

(a)



(b)

Fig. 1: *Sample training dataset of human detection framework, (a) Positive dataset and (b) negative dataset.*

Testing

True positive

False negative

The system detected (true) the image as human (positive).
Correctly identified

The system detected (false) the image as non-human (negative).
Incorrectly rejected

Testing (Cont.)

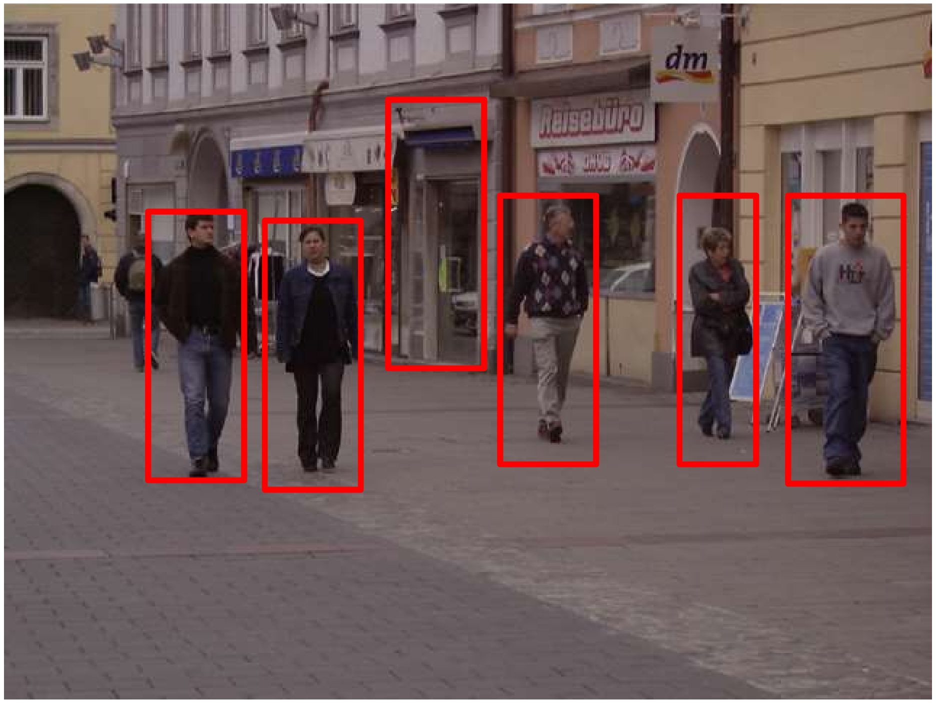
True negative

False positive

The system detected (true) the image as non-human (negative).
Correctly rejected

The system detected (false) the image as human (positive).
Incorrectly Identified

Precision and Recall



True positive = 5
False positive = 1
False negative=2

$$\text{Precision}=\frac{\text{True positive}}{\text{True positive + False positive}}=\frac{5}{6}$$

Precision is the fraction of retrieved documents that are relevant.

$$\text{Recall}=\frac{\text{True positive}}{\text{True positive + False negative}}=\frac{5}{7}$$

Recall is the fraction of the relevant documents that are successfully retrieved.

Precision & Recall

- Precision is the number of accurate positive prediction, divided by total number of positive predictions. It shows performance with respect to false positive.

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

- Recall is the number of accurate positive prediction, divided by total number of actual positive label.
- Recall is also called sensitivity, True Positive Rate (TPR). It shows performance with respect to false negative.

$$recall = \frac{TP}{TP+FN}$$

Specificity & Others

- Specificity is the number of accurate negative prediction, divided by total number of actual negative label. It is the opposite of Recall. It is also called True Negative Rate (TNR).

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

- Some other measures are –
  - False Negative Rate (FNR) / Miss rate

    $$Miss\ rate = 1 - Recall$$

  - False Positive Rate (FPR) / Fall out

    $$Fall\ out = 1 - Specificity$$

Other terminologies

- Miss rate = 1- recall

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

Accuracy is a metric that measures how often an ML or DL model correctly predicts the outcome. In other words, accuracy answers the question: how often the model is right?

- $$\text{F1 score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} = 2TP/(2TP+FP+FN)$$

The F1 score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

Cont.

- **Accuracy** is used when the True Positives and True negatives **are more important while** **F1-score** is used when the False Negatives and False Positives **are crucia l**. **Accuracy** can be used when the class distribution is similar while **F1-score** is a better metric when there are imbalanced classes.

Accuracy

- Accuracy is total number of correct predictions, both positive and negative, divided by the total number of samples.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

- Safe to use it when the dataset are nearly balanced over all target class, otherwise it may produce misleading result.

F-measure

- In order to get a average, balanced measure between precision and recall, F-measure is used. F1-score is the harmonic mean of precision and recall. It a special case of generalized F_β-measure.

$$F_1 score = 2 * \frac{precision * recall}{precision + recall}$$

$$F_\beta score = (1 + \beta^2) * \frac{precision * recall}{\beta^2 * precision + recall}$$

The value 2 & 0.5 for β has been also used in different research works.

ROC-AUC Curve

- The Receiver Operating Characteristic (ROC) curve is a graphical representation that shows the performance of the binary classifier based proposed method.
- To draw an ROC curve, False Positive Rate is plotted on *x* axis and Recall (True Positive Rate) is plotted on *y* axis.

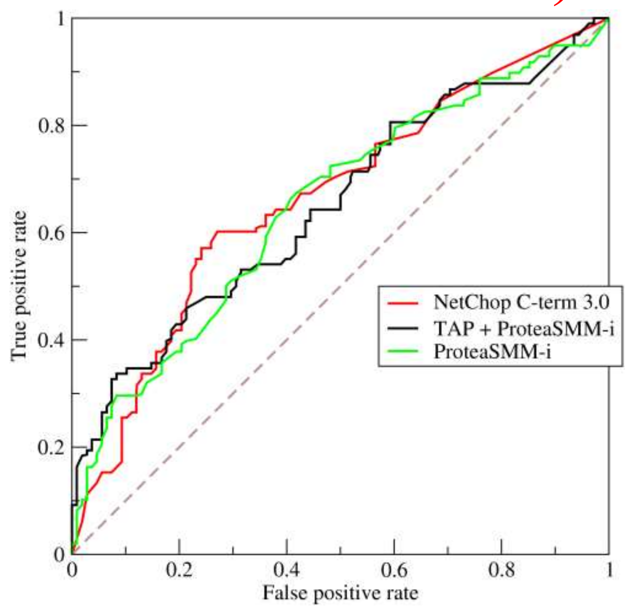
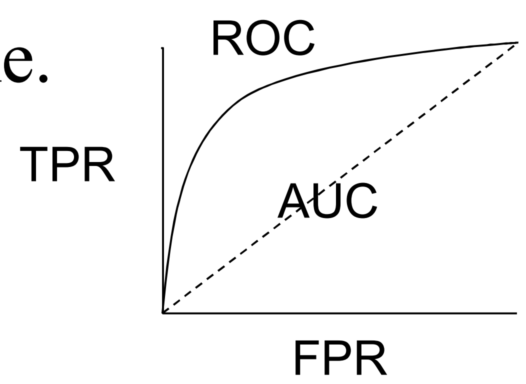


Fig. 2: *ROC curve of three predictors of peptide cleaving in the proteasome.*

ROC-AUC Curve

- Receive Operating Curve (ROC) is a probability curve that is plotted with TPR on the y-axis against FPR on the x-axis.
- Area Under The Curve (AUC) shows the performance of the model in case of classification. The higher the AUC, the better the model classifies between the class label. The range of AUC is [0,1). For multi-class classification , multiple ROC curve have to be drawn using the One vs. All scheme.



REFERENCES

1. INRIA Person dataset: http://pascal.inrialpes.fr/data/human/, accessed August 2017
2. https://en.wikipedia.org/wiki/Precision_and_recall/, accessed August 2017

Thank you all