Score Normalisation

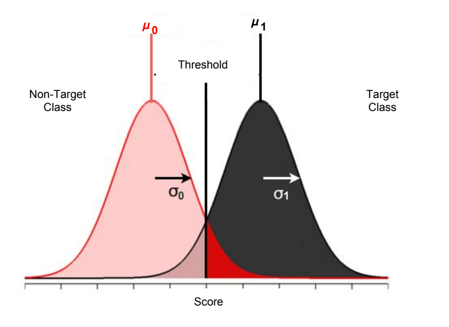
Pattern Recognition, Jul-Nov 2019

Indian Institute of Technology Madras

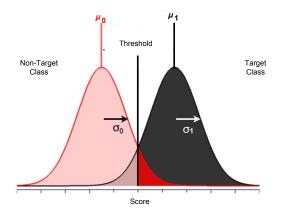
August 26, 2019

Target and Non-Target scores

- In any N class problem, the class which is of intrest can be consodired as target class and all the other classes can be pooled together and called as non-target class.
- The scores for the target and non target classses form a gaussian distibution as shown in the below figure.



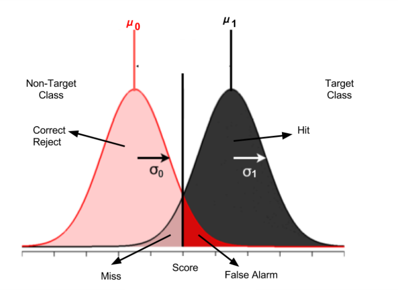
Confusion Matrix



Based on the thershold chossen on the score, every test example may result in one of the following four outcomes of confusion matrix.

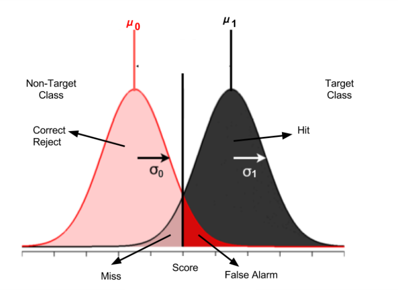
		Ground Truth	
		Target	Non- Target
Model Prediction	Target	HIT (True Positive)	False Alarm (False Positive)
	Non-Target	Miss (False Negative)	Correct Reject (True Negative)

Confusion Matrix



		Ground Truth	
		Target	Non- Target
Model Prediction	Target	HIT (True Positive)	False Alarm (False Positive)
	Non-Target	Miss (False Negative)	Correct Reject (True Negative)

Confusion Matrix



- For a ideal case, we would like the area under the curve to be minimum for "False Alarm" and "Miss". i.e., the means of the target and non-target scores should be far apart from each other.

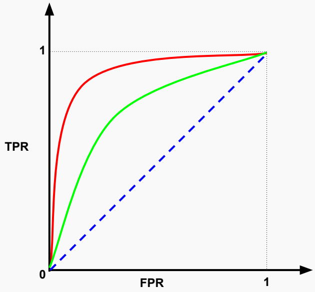
Messures of error from confusion matrix

- True postive rate or Recall $TPR = \frac{TP}{TP+FN}$
- False postive rate $FPR = \frac{FP}{FP+TN}$
- Accuracy $= \frac{TP+FP}{TP+FN+FP+TN}$
- Precission or positive predictive value$= \frac{TP}{TP+FP}$
- F Meassure $= 2\frac{precission \times recall}{precission + recall}$
- For other error meassures based on confusion matrix, please refer this wiki: https://en.wikipedia.org/wiki/Confusion_matrix

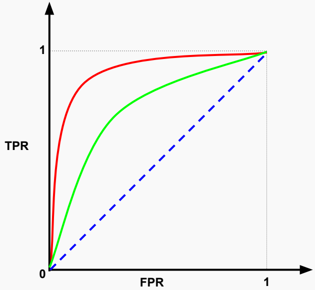
Receiver operating characteristic

Definition

In a ROC curve the true positive rate is plotted in function of the false positive rate for various decision thresold on the score



Receiver operating characteristic



In the above figure,

- The blue line denotes a random system.
- The red line denotes best system.

Detection error tradeoff curves (DET Curve)

- DET assumes that target and not target score follows a gaussian distribution

$$\text{True Negative} \sim \mathbb{N}(\mu_0, \sigma_0) \tag{1}$$
$$\text{True Positive} \sim \mathbb{N}(\mu_1, \sigma_1) \tag{2}$$
$$\tag{3}$$

- Therefore, the probality of miss(m) and false alarm(fa) can be written as:

$$P_m(t) = \int_{-\infty}^{t} e^{-\frac{(x-\mu_1)^2}{\sigma_1}} \quad dx = \phi\left(\frac{t-\mu_1}{\sigma_1}\right) \tag{4}$$
$$P_{fa}(t) = \int_{t}^{\infty} e^{-\frac{(x-\mu_0)^2}{\sigma_0}} \quad dx = \phi\left(\frac{\mu_0-t}{\sigma_0}\right) \tag{5}$$

Detection error tradeoff curves (DET Curve)

- Equation 4 and 5 can be rewritten as:

$$\frac{t-\mu_1}{\sigma_1} = \phi^{-1}\left(P_m(t)\right) \tag{6}$$

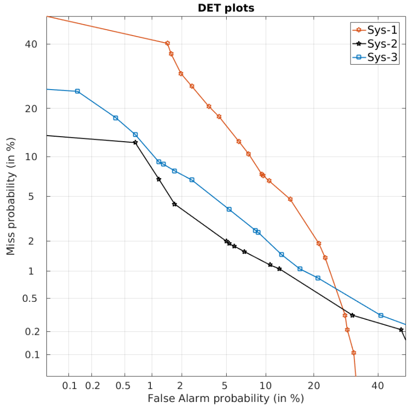
$$\frac{\mu_0 - t}{\sigma_1} = \phi^{-1}\left(P_{fa}(t)\right) \tag{7}$$

- Equating for t we get:

$$\phi^{-1}\left(P_m(t)\right) = \frac{-\sigma_0}{-\sigma_1}\phi^{-1}\left(P_{fa}(t)\right) + \frac{\mu_0-\mu_1}{\sigma_1} \tag{8}$$

- Equation 8 denotes the probabily of miss as a linear function of probality of false alaram. The plot of $\phi^{-1}\left(P_m(t)\right)$ vs $\phi^{-1}\left(P_{fa}(t)\right)$ is called as the DET curve.

Example of DET curve



Normalization using an impostor model

- Let $log(P(W_i|\theta))$ be the likelihood of the observation points θ belonging to the class w_i
- The normalized likelihood is given by

$$log(P(W_i|\theta)) = log(P(\theta|W_i)) - log(P(\theta|W_N))$$

  W_N : Cohort model
- In place of a cohort model, a world model can also be used
- A world model is trained using data from all the classes

Z-norm

- For each class W_i, a set of N cohort models are chosen
- These models are also called as Impostor models
- For each class, impostor mean (μ_I) and standard deviation (σ_I) are calculated
- For the test sample, the new normalized score is computed as

$$S = \frac{log(P(W_i|\theta)) - \mu_I}{\sigma_I}$$

T-norm

- Performing Z-norm is not adequate if there is a variability between the train and test
- In Test norm (T-norm), the mean and standard deviation are calculated during testing

$$S_T = \frac{log(P(W_i|\theta)) - \mu'_I}{\sigma'_I}$$