Comments on the ROC Curve

April 11, 2019

Bayesian Detection

- In Bayesian Detection problems, we form the likelihood ratio (or MAP test in the case of multiple simple hypotheses) and manipulate the ratio to form an easily computable test statistic.
- Next, we use the test statistic to compute P_D and P_{FA}.
- The main design criterion are the loss $L(0|1)$, $L(1|0)$ and sometimes the prior.
- In Frequentist detection, we don't have any prior information, so we can't form the Bayesian Risk as an objective function.
- One possible design criterion is to use P_{FA} to choose the threshold for the Likelihood Ratio Test.

Frequentist/Neyman-Pearson Process

- Choose an upper bound on the probability of false alarm α, for example $\alpha = 5\% = 0.05$.
- Formulate Likelihood Ratio $\Lambda(\boldsymbol{x}) = \frac{p(\boldsymbol{x};\theta_1)}{p(\boldsymbol{x};\theta_0)} \gtrless \lambda$.
- Find the critical value of the likelihood ratio λ by solving

$$\int_{\boldsymbol{x}:\Lambda(\boldsymbol{x})>\lambda} p(\boldsymbol{x};\theta_0)d\boldsymbol{x} = \alpha$$

  for λ.
- This can be difficult (i.e. solving integral equations, or for complicated likelihood functions).
- It's usually easier to fix a λ and calculate the probabilities of false alarm and detection, and then find the threshold corresponding to the desired α.

Receiver Operating Characteristic

- $P_D(\lambda) = \int_{\boldsymbol{x}:\Lambda(\boldsymbol{x})>\lambda} p(\boldsymbol{x};\theta_1)d\boldsymbol{x}$ - Plot the probability of detection as a function of λ on the y-axis.
- $P_{FA}(\lambda) = \int_{\boldsymbol{x}:\Lambda(\boldsymbol{x})>\lambda} p(\boldsymbol{x};\theta_0)d\boldsymbol{x}$ - Plot the probability of false alarm on the x-axis.
- This parametric curve is called the Receiver Operating Characteristic, whose name comes from radar.

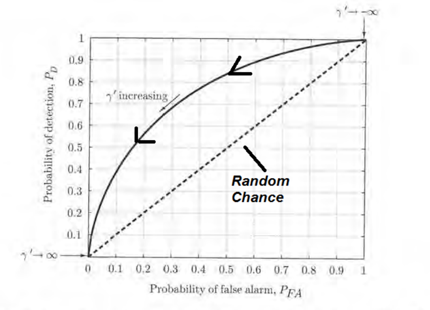

Figure 1: The ROC curve for the DC Level with Additive White Gaussian Noise Example. Note that as $\lambda = -\infty$, $P_D = P_F = 1$.

Desireable ROCs

- The goal is to maximize P_D while minimizing P_F.
- This means we want the ROC Curve to pull towards the top left of the plot.
- Curves below the random chance line have higher P_F than P_D.
- Area under the roc curve (AUC) is used as a metric to determine how "good" a ROC curve is.

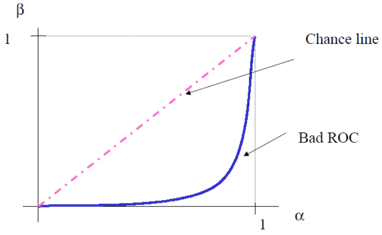

Figure 2: A poorly performing ROC curve.

ROC Curve, NP, and Bayesian Detection

- The Bayesian test sets $\lambda = \frac{\pi(H_0)L(0|1)}{\pi(H_1)L(1|0)}$
- The Neyman-Pearson Likelihood Ratio Test sets a specific P_F and then finds λ.
- In the example of Additive Gaussian White Noise, the Bayesian approach increase its probability of detection at the expense of false positives, but the NP approach generally results in very low P_D.

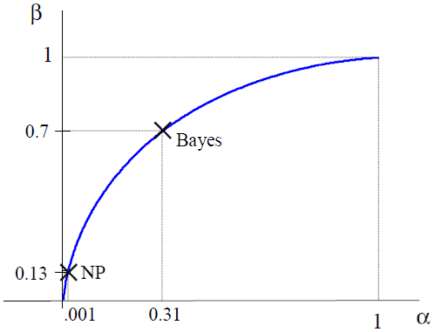

Figure 3: ROC Curve for Additive White Gaussian Noise example, here $\alpha = 0.001$

Connection to Machine Learning

- In binary classification problems, we are interested in deciding whether data belongs to a positive or negative class.
- Classification algorithms usually take the form of:

$$s(\mathbf{x}) \underset{\text{Neg}}{\overset{\text{Pos}}{\gtrless}} \lambda$$

  where $s(\cdot)$ is some score function that is higher for a positive and lower for a negative case.
- For example, $s(\mathbf{x}) = p(\theta = \theta_1|\mathbf{x})$.
- Some terms:
  - True Positive Rate (TPR): The proportion of positives correctly classified - analogous to P_D. - Also called sensitivity, recall, power.
  - False Positive Rate (FPR) : The proportion of negatives classified as positives - analogous to P_{FA} - Also called Type I Error.
  - True Negative Rate (TNR): The proportion of negatives correctly classified - analogous to $1 - P_{FA}$ - also called Specificity.
  - False Negative Rate (FNR): The proportion of positives incorrectly classified as negaties - analogous to $P_{Miss} = 1 - P_D$ - Also called Type II Error.

Connection to Machine Learning Cont.

- The difference in Machine Learning is that the quantities in the previous slide are empirical.
- They are calculated using a test or validation set, which consists of data not used to train the model, and things like the TPR and FPR are calculated by counting correct/incorrect classifications.
- As $N \to \infty$ this approach lines up with the probabilistic methods.
- We pick the threshold based on the quantity that is most important to the problem, or for a more balanced approach we minimize $(1 - P_D)^2 + P_{FA}^2$.
- This picks the point closest to the top left of the ROC.
- Other ways to choose a threshold include the precision-recall curve, which is more focused on the positives than the negatives, or accuracy, which is a weighted average of TPR and FPR.