Detection Theory. Non-bayesian binary detection. Evaluation of classifiers.

A Modern Theory of Detection and Estimation.
Block-2: Analytical Detection

Emilio Parrado-Hernández, emilio.parrado@uc3m.es

October 27, 2022

Index

Index

LRT rule

Binary deciders are defined by

- Discriminant function $g(\mathbf{x})$
- Threshold η

LRT tests

The discriminant function is the likelihood ratio

$$g(\mathbf{x}) = \frac{p_{\mathbf{X}|H}(\mathbf{x}|1)}{p_{\mathbf{X}|H}(\mathbf{x}|0)} = \lambda \underset{D=0}{\overset{D=1}{\gtrless}} \eta$$

Example

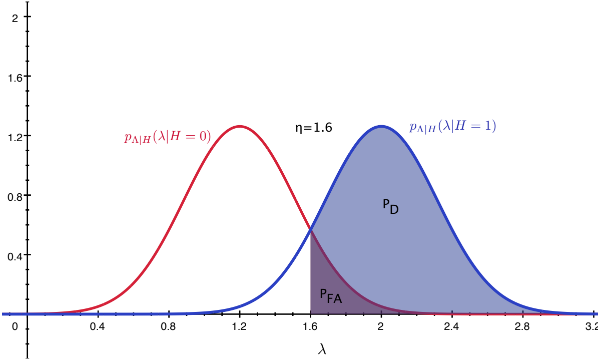
- ML: $\eta = 1$
- MAP: $\eta = \frac{P_H(0)}{P_H(1)}$
- Bayesian: $\eta = \frac{P_H(0)}{P_H(1)} \frac{c_{10} - c_{00}}{c_{01} - c_{11}}$

LRT decider depending on parameter η

Every possible value of $\eta \in [-\infty, \infty]$ originates a [potentially] different LRT.
The value of η determines the performance of each LRT decider:

$$P_{\mathrm{FA}} = \int_{\eta}^{\infty} p_{\Lambda|0}(\lambda|0) d\lambda$$

$$P_D = \int_{\eta}^{\infty} p_{\Lambda|1}(\lambda|1) d\lambda$$

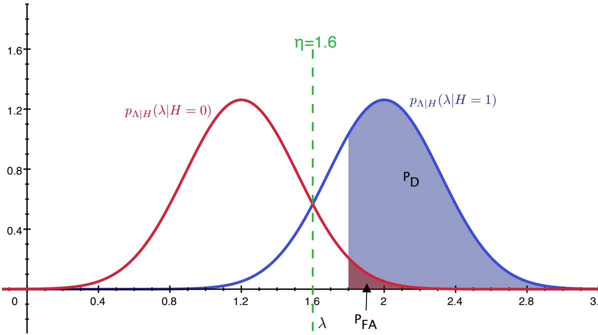


LRT decider depending on parameter η

Every possible value of $\eta \in [-\infty, \infty]$ originates a [potentially] different LRT. The value of η determines the performance of each LRT decider:

$$P_{\mathrm{FA}} = \int_{\eta}^{\infty} p_{\Lambda|0}(\lambda|0)d\lambda$$

$$P_D = \int_{\eta}^{\infty} p_{\Lambda|1}(\lambda|1)d\lambda$$

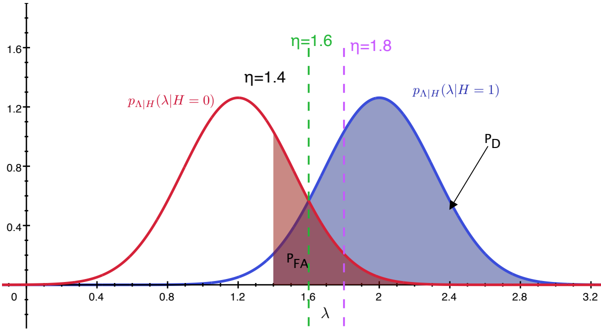


LRT decider depending on parameter η

Every possible value of $\eta \in [-\infty, \infty]$ originates a [potentially] different LRT. The value of η determines the performance of each LRT decider:

$$P_{\mathrm{FA}} = \int_{\eta}^{\infty} p_{\Lambda|0}(\lambda|0)d\lambda$$

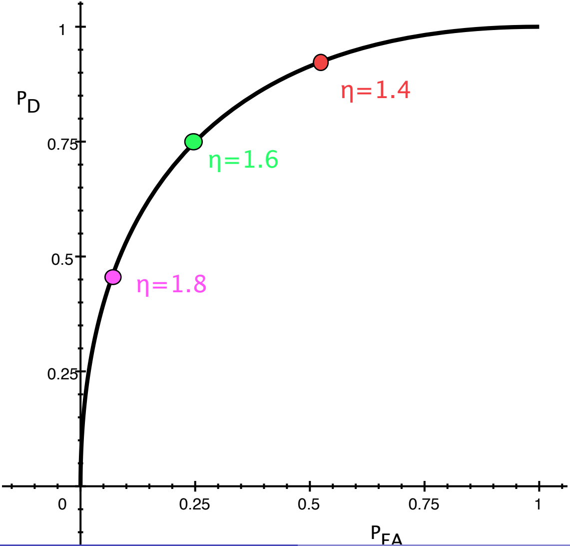
$$P_D = \int_{\eta}^{\infty} p_{\Lambda|1}(\lambda|1)d\lambda$$



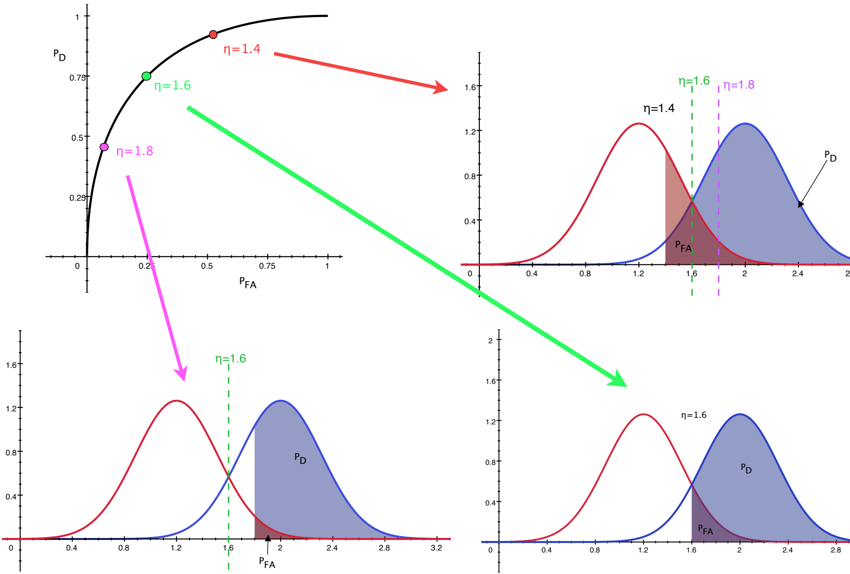
Receiver Operating Characteristic curve

Assigning values to parameter η we can tune a range of deciders with different performances in terms of P_{FA} and P_D.

The ROC curve is a plot of P_D vs. P_{FA} indexed by the different values of η



Receiver Operating Characteristic curve



Index

Example

A binary problem with likelihoods

$$p_{X|H}(x|1) = 2x, \qquad 0 \leq x \leq 1$$
$$p_{X|H}(x|0) = 2(1-x), \qquad 0 \leq x \leq 1$$

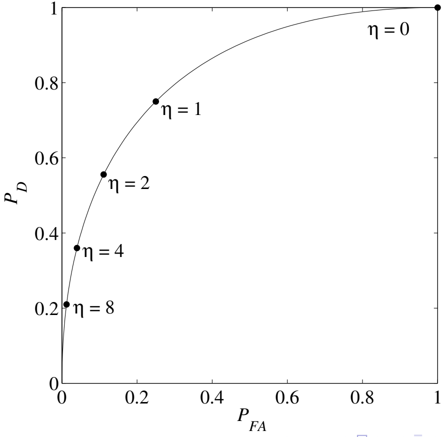
yields the following LRT

$$\frac{x}{1-x} \underset{D=0}{\overset{D=1}{\gtrless}} \eta$$

with performances

$$P_{\mathrm{FA}}(\eta) = \frac{1}{(1+\eta)^2}$$

$$P_{\mathrm{D}} = 1 - \frac{\eta^2}{(1+\eta)^2}$$



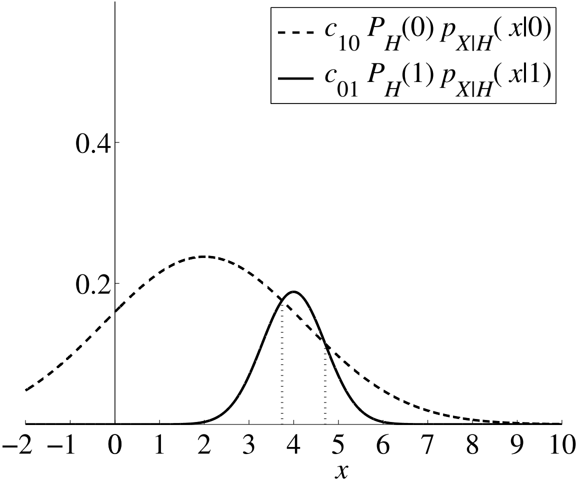
Index

Another example (from past lecture)

Consider a binary decision problem with Gaussian likelihoods with means $m_0 = 2$ and $m_1 = 4$ and variances $v_0 = 5$ and $v_1 = 0.5$

And Priors $P_H(0) = 2/3$ and $P_H(1) = 1/3$

And costs $c_{10} = 2$, $c_{01} = 1$ and $c_{00} = c_{11} = 0$



From LRT to g(x)

It is a curve parameterized by η in the LRT

$$\frac{p_{\mathbf{x}|H}(\mathbf{x}|1)}{p_{\mathbf{x}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\gtrless}} \eta$$

After some math we have arrived at

$$-9x^2 + 76x \underset{D=0}{\overset{D=1}{\gtrless}} 156 + 10\log(\frac{\eta}{\sqrt{10}})$$

We can rewrite the decider using a new threshold $\eta' = 156 + 10\log(\frac{\eta}{\sqrt{10}})$

$$-9x^2 + 76x \underset{D=0}{\overset{D=1}{\gtrless}} \eta'$$

Notice a quadratic decider involves two thresholds on the observation!!

Computing $P_{\mathrm{FA}}(\eta)$ and $P_{\mathrm{D}}(\eta)$

1. Selecting η determines $\eta'(\eta) = 156 + 10\log(\frac{\eta}{\sqrt{10}})$
2. $\eta'(\eta)$ determines two thresholds on the observation

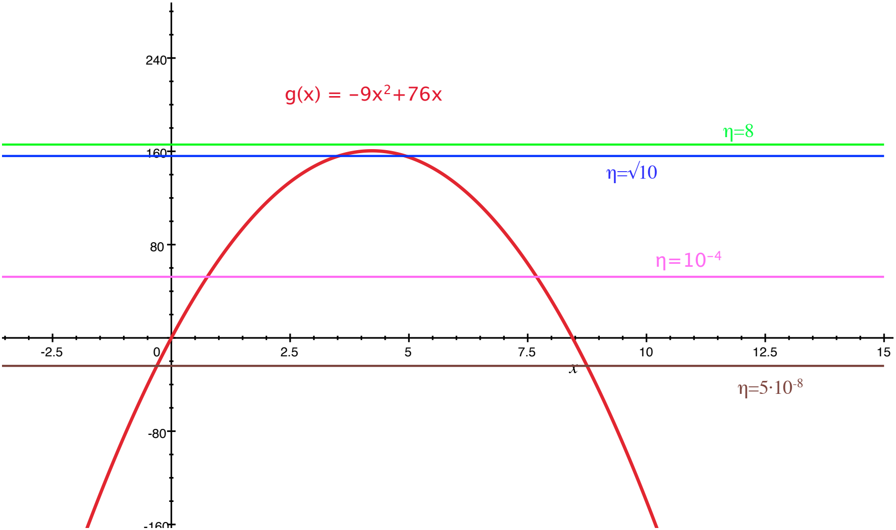
$$x_1(\eta') = \frac{-76 - \sqrt{76^2 + 4(-9)\eta'}}{2(-9)}$$

$$x_2(\eta') = \frac{-76 + \sqrt{76^2 + 4(-9)\eta'}}{2(-9)}$$

3. $$P_{\mathrm{FA}}(\eta) = \int_{x_1(\eta'(\eta))}^{x_2(\eta'(\eta))} p_{x|0}(x|0)dx$$

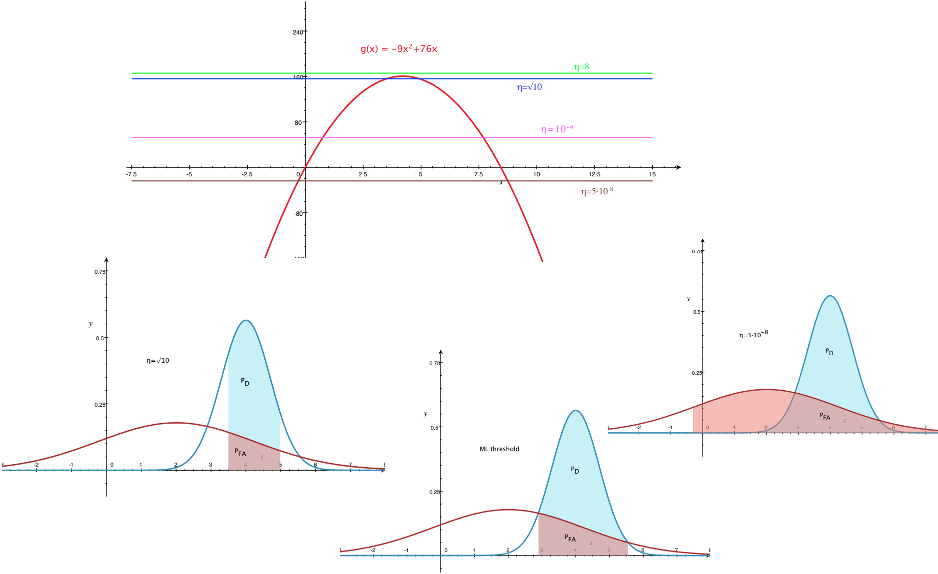
4. $$P_{\mathrm{D}}(\eta) = \int_{x_1(\eta'(\eta))}^{x_2(\eta'(\eta))} p_{x|1}(x|1)dx$$

LRT



240
160
80
-80
-160
-2.5
0
2.5
5
7.5
10
12.5
15
x
g(x) = −9x²+76x
η=8
η=√10
η=10⁻⁴
η=5·10⁻⁸

LRT, P_{FA} y P_{D}



Complementary error function erfc(x)

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$$

$$P_{\text{FA}} = \frac{1}{\sqrt{10\pi}} \int_{x_1}^{x_2} \exp\left(-\frac{(x-2)^2}{10}\right) dx$$

$$= \frac{1}{\sqrt{10\pi}} \left(\int_{x_1}^{\infty} \exp\left(-\frac{(x-2)^2}{10}\right) dx - \int_{x_2}^{\infty} \exp\left(-\frac{(x-2)^2}{10}\right) dx \right)$$

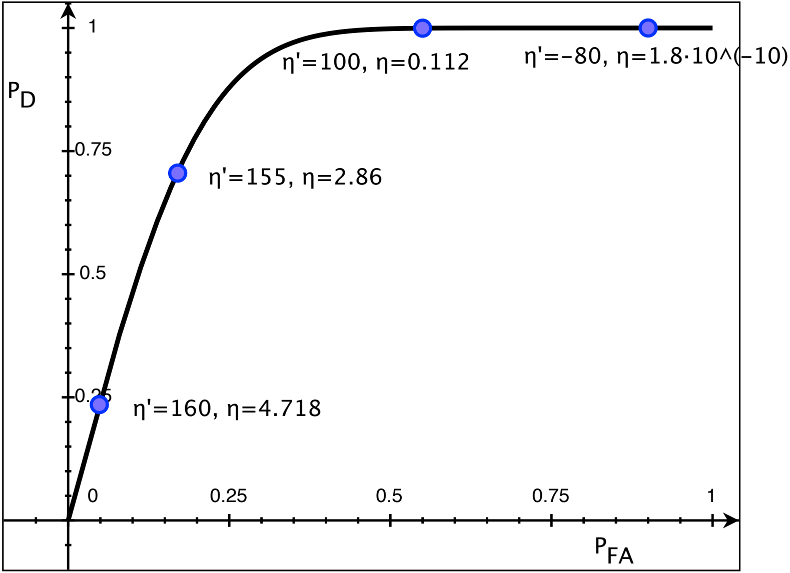
Change of variable $t = \frac{(x-2)}{\sqrt{10}}$

$$P_{\text{FA}} = \frac{1}{\sqrt{10\pi}} \left(\int_{\frac{(x_1-2)}{\sqrt{10}}}^{\infty} \exp(-t^2)\sqrt{10}dt - \int_{\frac{(x_2-2)}{\sqrt{10}}}^{\infty} \exp(-t^2)\sqrt{10}dt \right)$$

$$P_{\text{FA}}(\eta) = 0.5 \left[\text{erfc}\left(\frac{(x_1(\eta)-2)}{\sqrt{10}} \right) - \text{erfc}\left(\frac{(x_2(\eta)-2)}{\sqrt{10}} \right) \right]$$

Analogously we can compute $P_{\text{D}}(\eta) = 0.5\left[\text{erfc}(x_1(\eta)-4) - \text{erfc}(x_2(\eta)-4)\right]$

ROC curve


1
0.75
0.5
0.25
0
0.25
0.5
0.75
1
P_D
P_FA
η'=100, η=0.112
η'=-80, η=1.8·10^(-10)
η'=155, η=2.86
η'=160, η=4.718

Index

ROC of a generic discriminant function

Up to now we have introduced the ROC curve as a means to evaluate the performance of an LRT.

However, in essence, to compute a ROC we only need P_{FA} and P_{D}. And these two quantities can be calculated for any given binary classifier.

Therefore the ROC curve can be drawn for any binary classifier even if its discriminant function does not arise from an LRT:

- Express the classifier as a discriminant function of the observations $f(\mathbf{x})$ and a threshold η
- Express P_{FA} and P_{D} as functions of η
- Give values to η and draw $P_{\mathrm{FA}}(\eta)$ vs. $P_{\mathrm{D}}(\eta)$

ROC curves to compare the capabilities of discriminant functions

Each discriminant function $f(\mathbf{x})$ can be represented by its associated ROC curve (the one that can be drawn varying η freely).

A perfect classifier would be able to perform in the ROC curve point $(P_{\mathrm{FA}} = 0.0, P_{\mathrm{D}} = 1.0)$. That is, without errors.

A given discriminant function is able implement a perfect classifier as long as its corresponding ROC curve includes the point $(P_{\mathrm{FA}} = 0.0, P_{\mathrm{D}} = 1.0)$.

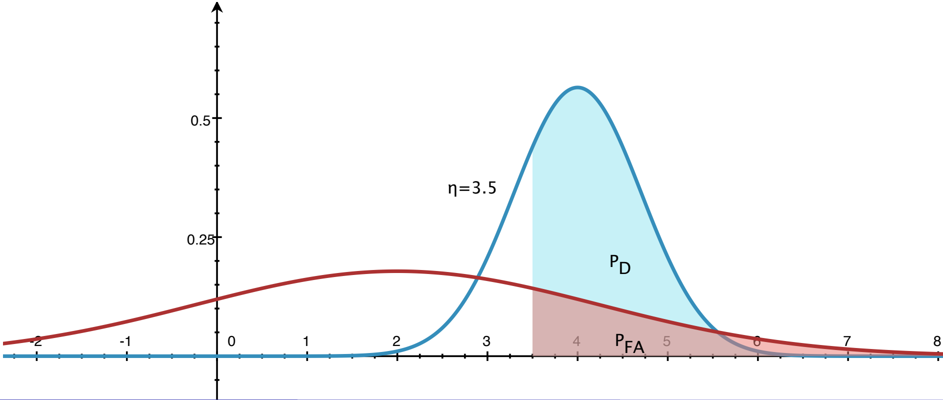
We can use the distance between a ROC curve and point $(P_{\mathrm{FA}} = 0.0, P_{\mathrm{D}} = 1.0)$ as quality measure for a discriminant function (the closer the ROC curve is from $(P_{\mathrm{FA}} = 0.0, P_{\mathrm{D}} = 1.0)$ the better the classifier).

Remember than the ROC curve corresponding to the LRT includes the Bayes classifier. Since the Bayes classifier is optimum, the LRT ROC curve imposes a bound on the performance achievable in a binary problem: there can't be ROC curves between the LRT ROC curve and $(P_{\mathrm{FA}} = 0.0, P_{\mathrm{D}} = 1.0)$

Example: Single threshold classifier in the previous Gaussian case

Let's analyse the performance of a classifier consisting in a threshold on the observations on the previous problem. The classifier is

$$x \underset{D=0}{\overset{D=1}{\gtrless}} \mu$$



Computing $P_{\mathrm{FA}}(\eta)$ and $P_{\mathrm{D}}(\eta)$

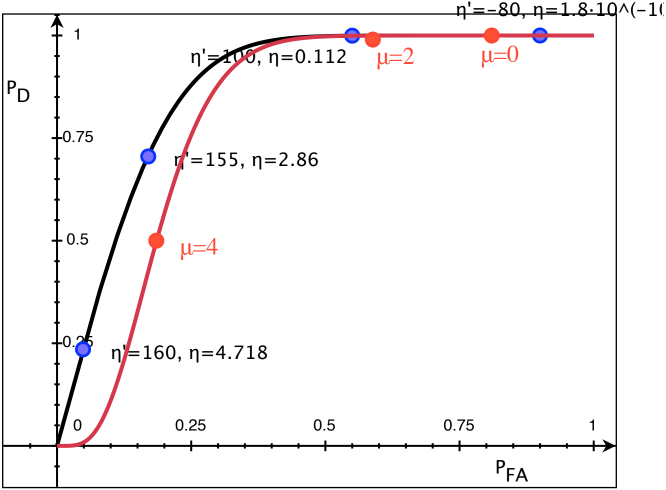
1. μ determines two decision regions with a single threshold
2. $$P_{\mathrm{FA}}(\mu) = \int_{\mu}^{\infty} p_{x|0}(x|0)dx$$
3. $$P_{\mathrm{D}}(\mu) = \int_{\mu}^{\infty} p_{x|1}(x|1)dx$$

Using $\mathrm{erfc}(x)$

$$P_{\mathrm{FA}}(\mu) = 0.5\mathrm{erfc}\left(\frac{\mu - 2}{\sqrt{10}}\right)$$

$$P_{\mathrm{D}}(\mu) = 0.5\mathrm{erfc}\left(\mu - 4\right)$$

ROC for the single threshold classifier



Black: ROC of the LRT. Red: ROC of the single threshold classifier

Index

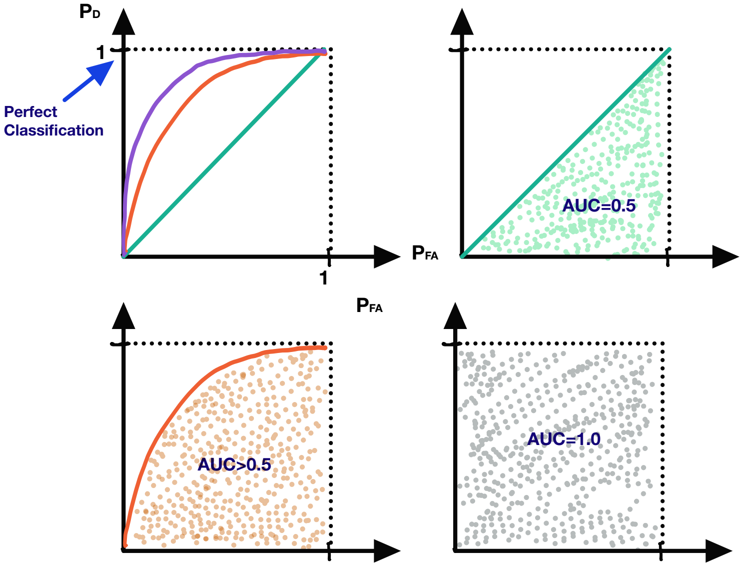
Index

Area under the ROC curve

We can asses the quality of a binary classifier by evaluating the distance between its ROC curve and the **perfect performance** working point: $(P_{\mathrm{FA}} = 0, P_D = 1)$
However, the quantification of this feature can result tricky.
An alternative is to evaluate the **Area under the ROC curve** as proxy to the quality of a detector:

- It's is bounded between 1 (perfect classification) and 0 (misguided classifier)
- Notice however the worse possible case is when AUC=0.5: A classifier that is *always wrong* is as perfect as a classifier that is *always right*, you just need to negate its output.

AUC for binary classifiers



Index

Sensitivity and Specificity

Sensitivity

Probability of detecting correctly the presence of a target

$$\text{Sensitivity} \; = P\{D = 1|H = 1\} \int_{\eta}^{\infty} p_{\Lambda|H}(\lambda|H = 1)d\lambda$$

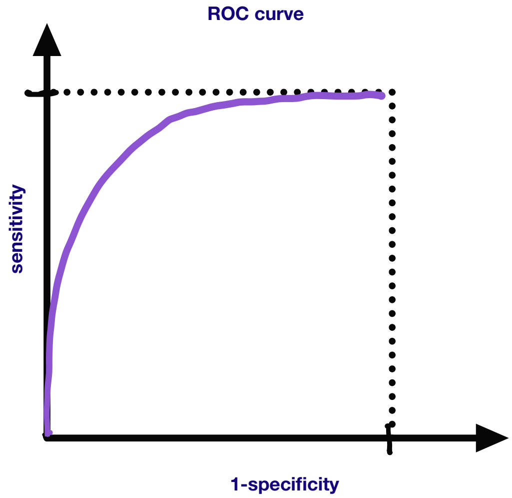
Specificity

Probability of detecting correctly the absence of a target

$$\text{Specificity} \; = P\{D = 0|H = 0\} \int_{-\infty}^{\eta} p_{\Lambda|H}(\lambda|H = 0)d\lambda$$

Sensitivity and Specificity and ROC curve

- Sensitivity is the Probability of Detection
- Specificity $= 1 - P_{\text{FA}}$



Index

Index

Neyman-Pearson Detectors

In some cases the importance of falses alarms is so great that the probability of false alarm becomes a critical criterion in the design of the classifier

Neyman-Pearson

$$\phi^* = \arg\max_{\phi}\{P_{\mathrm{D}}\} \text{ subject to } P_{\mathrm{FA}} \leq \alpha$$

- Impose a bound on P_{FA}.
- Maximize P_{D} but guaranteeing that P_{FA} stays below the bound.

Procedure to determine an LRT Neyman Pearson Detector:

1. Obtain the LRT as a function of η
2. Express P_{FA} as a function of η
3. Make $P_{\mathrm{FA}} = \alpha$ and solve for η
4. Plug that value of η in the LRT

Example

Determine the LRT Neyman-Pearson classifier with $P_{\mathrm{FA}} \leq \alpha = 0.1$ in a binary problem with likelihoods

$$p_{X|H}(x|1) = 2x, \qquad 0 \leq x \leq 1$$
$$p_{X|H}(x|0) = 2(1-x), \qquad 0 \leq x \leq 1$$

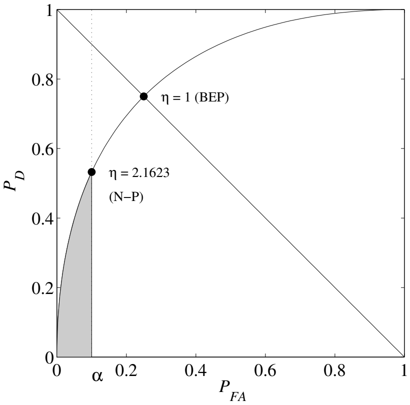
The LRT is $\dfrac{x}{1-x} \underset{D=0}{\overset{D=1}{\gtrless}} \eta$

with $P_{\mathrm{FA}}(\eta) = \dfrac{1}{(1+\eta)^2}$

Making $P_{\mathrm{FA}} = \alpha = 0.1$:

$$\eta = \frac{1}{\sqrt{P_{\mathrm{FA}}}} - 1 \approx 2.1623$$

and $P_{\mathrm{D}} = 1 - \dfrac{\eta^2}{(1+\eta)^2} \approx 0.53$



Index

Minimax strategy

Minimax is a broadly used strategy to take decisions. It takes the decision that minimizes the effects of the worst case scenario.
Example: Imagine the following cost policy (no likelihoods available) (remember c_{dh} is the cost of deciding d when the right hypothesis is h)

$$C = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 1 & 5 \\ 6 & 1 & 0 \end{bmatrix}$$

The worst case scenarios for each decision are:

$$\begin{bmatrix} c_{02} = 4 \\ c_{12} = 5 \\ c_{03} = 6 \end{bmatrix}$$

So the minimax decision is $D = 0$. Notice that $D = 2$ would eventually lead to $c = 0$ if $H = 2$, but to $c = 6$ if the worst case $H = 0$ happens

Minimax classifier

The extension of the minimax strategy to a classification problem defined in terms of P_{FA} and P_{D} becomes

Minimax binary classifier

Choose η so that $P_{\mathrm{FA}} = P_{\mathrm{M}} = 1 - P_{\mathrm{D}}$

This point is the intersection of the ROC curve with the line $P_{\mathrm{FA}} = 1 - P_{\mathrm{D}}$. We term it **Break Even Point** (BEP).

It is easy to show that for this classifier $P_e = P_{\mathrm{FA}} = P_{\mathrm{M}}$, what means that P_e in this case is independent of the prior probabilities of the hypotheses. Therefore this classifier is robust to changes in the prior probabilities of the classes.