UE Machine Learning
Binary classifier & statistical hypothesis testing
Lab. 3 (Part II)
The Neyman-Pearson test and the GLRT

dominique.pastor@imt-atlantique.fr

Introduction

Maximum Likelihood Estimate (MLE). Suppse that X is a random vector that has a pdf depending on some parameter θ. We denote this pdf $p(x;\theta)$ when x is a realisation of X. The maximum likelihood estimate (MLE) of θ when we have the realization x of X is then defined as the value $\widehat{\theta}_p(x)$ that maximizes $p(x;\theta)$. We thus have

$$\widehat{\theta}(x;p) \stackrel{\text{def}}{=} \arg_\theta \max p(x;\theta)$$

The notation used above for the MLE specifies the dependency of this estimate with the pdf p. Indeed, two pdfs p and p' may depend on the same parameter θ, whereas the MLEs $\widehat{\theta}(x;p)$ and $\widehat{\theta}(x;p')$ may differ. However, this dependency is generally dropped for the sake of simplifying the notation. But, it must be kept in mind.

Generalized Likelihood Ratio Test (GLRT). When we write a statistical hypothesis testing problem in the form:

$$\begin{cases} \mathcal{H}_0 : X \sim p_0 \\ \mathcal{H}_1 : X \sim p_1 \end{cases} \tag{1}$$

we assume that the pdfs p_0 and p_1 are perfectly known. But it may happen that these pdfs depend on some unknown parameters. So, suppose that p_0 depend on θ_0 and p_1 depends on θ_1. The parameters θ_0 and θ_1 can be vectors. To specify these dependencies, let us denote by $p_i(x;\theta_i)$ the pdf p_i for $i = 0,1$ when the realisation of X is x. If the parameters are unknown, we cannot apply the Neyman-Pearson Lemma. The idea is then to replace all the unknown parameters in p_0 and p_1 by their maximum likelihood estimates in the expression of the likelihood ratio.

When we replace θ_0 and θ_1 by MLEs $\widehat{\theta_0}(x,p_0) = \arg_\theta \max p_0(x;\theta_0)$ and $\widehat{\theta_1}(x,p_1)$ in the likelihood ratio $L(x) = \dfrac{p_1(x;\theta_1)}{p_0(x;\theta_0)}$ for a given realisation x of X, we then obtain the Generalized Likelihood Ratio, which turns out to be equal to:

$$L_G(x) = \frac{p_1(x;\widehat{\theta_1}(x,p_1))}{p_0(x;\widehat{\theta_0}(x,p_0))} = \frac{\max_{\theta_1} p_1(x;\theta_1)}{\max_{\theta_0} p_0(x;\theta_0)}$$

It's the second equality we generally use to compute the generalized likelihood ratio. Indeed, we don't actually need the MLEs to derive the generalized likelihood ratio because only the values of the pdfs for these MLEs actually count to obtain it. Take a while to ponder over that!

- θ_0 and θ_1 may be the same parameter. But their estimate may differ under each hypothesis!
- What is known need not to be estimated! For instance, if θ_0 is known in $p_0(x;\theta_0)$ (and we often write $p_0(x)$ instead of $p_0(x;\theta_0)$ to mean that θ_0 is actually known), the generalized likelihood ratio reduces to: $L_G(x) = \max_{\theta_1} p_1(x;\theta_1)/p_0(x)$

The GLRT then involves comparing $L_G(x)$ to a threshold that we must choose in order to guarantee a specified probability of false alarm.

In general, the GLRT does not derive from a given optimality criterion. It's just a trick, a nice trick that achieves good results, as shown by the following exercise. In some cases, it can however be proved to be optimal with respect to specific critria for optimality, generally in connection with the invariance properties of the problem.

Problem

$\|\cdot\|$ designates the Eucilidean norm in $\mathbb{R}^N$. It is defined for any $y = (y_1, \ldots, y_N)^{\mathrm{T}} \in \mathbb{R}^N$ by $\|y\| = \sqrt{\sum_{n=1}^{N} y_n^2}$.

Let Y be a N-dimensional real random vector. We make the the following two hypotheses on Y:

$$\begin{cases} \mathcal{H}_0: & Y \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N) \\ \mathcal{H}_1: & Y \sim \mathcal{N}(A\xi_0, \sigma^2 \mathbf{I}_N) \end{cases}$$

where $\sigma \neq 0$ and $\xi_0 \in \mathbb{R}^N$ are known and $\mathbf{I}_N$ is the identity matrix with size $N \times N$. We assume that $\|\xi_0\| = 1$. Let $\alpha \in]0,1[$.

1) Calculate the Neyman-Pearson (NP) test with size α, assuming that A is known (the answer is given in the course). This test is said to be clairvoyant because A is assumed to be known. What is the power of this test depending on whether $A > 0$ or $A < 0$?

2) We now assume that A is unknown. The purpose is to calculate the GLRT and to compare it to the clairvoyant test.

1. Compute the maximum likelihood estimate $\widehat{A}_{\mathrm{MLE}}$ of A under $\mathcal{H}_1$.
2. Derive the GLRT with size α for testing $\mathcal{H}_0$ against $\mathcal{H}_1$. Comment the structure of this test.
3. Compute the power of this test and check out your result by Monte-Carlo simulations.
4. Compare the performance of the GLRT to that of the clairvoyant test, theoretically and by simulations. What are your conclusions?