

Assignment 2: Bayesian Decision Theory and Parametric Estimation

Submission: Tuesday September 8th
Groups of maximum 2 students

Prof. Fabio A. González
Machine Learning - 2015-II
Maestría en Ing. de Sistemas y Computación

-
- (2.5) Download the dataset from [this link](#). The dataset is a text file with a number of data samples, one per line. Each line has the following structure:

$$x_i y_i C_i,$$

where $(x_i, y_i) \in \mathbb{R}^2$ and $C_i \in \{0, 1, 2\}$.

- Use the data for classes 1 and 2 to estimate the parameters of a bivariate Gaussian distribution for each class. Assume that the covariance matrix is the same for both classes.
 - Write the parameters of the probability distribution functions for both classes.
 - Write a function that calculates the discriminant function for each class.
 - Draw a plot, where the regions corresponding to the different classes are shown with different colors. A region corresponding to a class is the set of points where the particular class discriminant function is maximum (decision regions, [Alp10] Sect. 3.4).
 - The boundary between both class regions must be a line. Calculate the equation of this line clearly explaining the deduction process. Draw the line along with the regions.
 - What happens with the boundary line if we change the prior probabilities of the classes? Illustrate with a graphical example.
- (1.0) Repeat steps (a) to (d) from previous item, but this time:
 - Use data from the three classes.
 - Estimate a different covariance matrix for each class.
 - (1.5) Repeat the previous item, but this time:
 - Use only a portion of the dataset (80% of the samples) to estimate the parameters of the probability distribution functions of each class.
 - Write a function that calculates the discriminant function for each class, taking into account the possibility of rejection with a cost λ and cost 1 for misclassification ([Alp10] Eq. (3.10)).
 - Classify the rest of the dataset that was not used for estimation (20%), using a classifier based on the discriminant functions. Evaluate the results using a confusion matrix.
 - The assignment must be submitted as an [IPython notebook](#) through the following [Dropbox file request](#), before midnight of the deadline date. The file must be named as `ml-assign2-unalusername1-unalusername2.ipynb`, where `unalusername` is the user name assigned by the university (include the usernames of all the members of the group).

References

- [Alp10] Alpaydin, E. 2010 Introduction to Machine Learning, 2Ed. The MIT Press.
- [DHS00] Duda, R. O., Hart, P. E., and Stork, D. G. 2000 Pattern Classification (2nd Edition). Wiley-Interscience.