

Bayesian Classifier

Assignment 1

Prof.	Panos Trahanias	trahania@csd.uoc.gr
T.A.	Myrto Vilia	mvillia@ics.forth.gr
T.A.	Despina - Ekaterini Argiropoulos	despary@csd.uoc.gr
T.A.	Michalis Savorianakis	csdp1387@csd.uoc.gr

Deadline Sunday 23:59, 16th of March 2025

Abstract

In this assignment, the task is to familiarize with Bayesian Decision Theory. Assignment's 1 task is the implementation of a Bayesian classifier and its evaluation on the given Heart Disease dataset.

Dataset: It is included in the *dataset.csv* file and it consists of 4 columns: 'trestbps', 'thalach', 'chol' and 'target'. Notice, the original dataset has 14 features in total, but only 3 are given for this Assignment. These columns represent measurements of resting blood pressure (trestbps), maximum heart rate achieved (thalach), and serum cholesterol (chol) for various individuals, except for the last one which denotes whether the individual has a heart disease (0) or not (1). Finally in the *parameters.txt* file the mean and covariance matrices are included, which are necessary for your implementation.

Note: This is a personal assignment and should be pursued individually and without use of any computerised AI facilities. Note that this will be automatically checked for every delivered assignment and for cases of relevant detections the whole assignment will be dropped. The assignment should be implemented entirely in Google Colaboratory following the delivered instructions below. **Comments regarding the submitted code are mandatory.**

Question A: The model (50/100)

1. Implement a function that computes the prior probabilities for each class (with and without heart disease).
2. Compute the posterior probabilities for each sample.
 - (a) Implement a function that takes as input one sample (one point in the 3-dimensional space), a mean and a covariance matrix and returns the likelihood of that point belonging to the class whose distribution is described by the given mean and covariance matrix.
 - (b) Implement a function that computes the posterior probabilities for each sample.
3. Print the predicted class for each sample, according to the posterior probabilities and then, compute the accuracy of your classifier.

Question B: Conditional Risk (30/100)

1. Compute the conditional risk for every sample.
2. Print the predicted class for each sample, in order to minimize the conditional risk. What is the accuracy now? Comment on what do you notice and why.

Question C: Plot Results (20/100)

1. Create a 2D scatter plot using only the features ['trestbps', 'thalach'], showing the distribution of each class.
2. Additionally, plot the decision boundary of your classifier.
3. Analyze the plot results. Does the data appear to be easily separable? Provide possible explanations for your observations.

* 5% Bonus for detailed comments

Deliverable

This assignment should be implemented entirely in Google Colaboratory. Google's notebook allows you to combine executable Python scripts with rich text in a single document. Your deliverable should be a single .ipynb file along with its corresponding .py file (both can be easily exported from Google Colaboratory). Every single question should be implemented in a single code block. Code blocks should be clearly and shortly explained (you may use the text boxes for that goal). Use **only** library functions for matrix operations and plots.