TP – Bayes

M.El Kamali, Q.Meteier,
L. Angelini, E. Mugellini,

Goals

During this practical work, you will practice with Bayes formula and create your own Gaussian Naive Bayes classifier.

1 Question 1: Titanic

- Use the Jupyter notebook "`titanic_bayes_stud.ipynb`" and get the Titanic dataset from Moodle.
- Answer the following question: "What are my chances of dying on the Titanic given my gender?"
    - Try to formulate the question with mathematical notations according to the Bayes law.
    - Complete the notebook.

2 Question 2: Bayesian classifier

- Use the Jupyter notebook "`TP Bayes.ipynb`" available on Moodle.
- Follow the instructions and add your code.
- Answer the questions.
- Write a summary with your conclusions and observations.

3 Question 3: post-tuning a trained model (optional)

Assume that you have trained a classifier for determining the gender (F=female,M=male) from images of human faces. The model provides posterior probabilities P($F|x$) and P($M|x$) of recognizing a female (F) or a male (M) given the image (x). Assume that your training data consists of 90% females and 10% males.

a. Let's assume you have a very blurred image x at the input of your system. The image is so blurred you have difficulties yourself to recognize if the person is a male or a female. What would be a reasonable output to the system? Provide numerical answers and explanations.
b. Let's now assume your model needs to be used in an environment with equi-probable distribution of females and males. In other words, the system should recognize females or males without making an "a priori" assumption on the gender of the user. Explain in detail how you need to correct the model output.

4 Deliverables

- Report with answers to all questions as well as your observations. It can be written directly in the Jupyter notebook
- Add to the archive the executed `.ipynb` file(s)

5 References

- https://docs.python.org/3/tutorial/datastructures.html
- https://docs.python.org/3/library/functions.html#zip
- https://docs.scipy.org/doc/numpy/reference/generated/numpy.mean.html
- https://docs.scipy.org/doc/numpy/reference/generated/numpy.std.html
- https://docs.scipy.org/doc/numpy/reference/generated/numpy.where.html
- http://matplotlib.org/examples/statistics/histogram_demo_features.html