

Homework 5

General guidelines:

- The homework is composed of two parts: Theoretical exercise and programming exercise.
- The theoretical exercise is written below, and its solution should be submitted as a pdf named as "HW#-<id>.pdf". For example: HW1-123456789.pdf
- All details on the programming exercise would usually appear only in a Jupyter notebook. Some general instructions might appear after the theoretical exercise, so please check it before jumping to the notebook.
- The answered notebook should be named similarly to the name of the pdf: "HW#-<id>.ipynb".
- Two files should be submitted in total: a pdf for the theoretical exercise, and a notebook for the programming exercise.'

Theoretical Exercise:

Note that the density function of the normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Bayes Classifier

1. Denote a classification problem with 1 feature and binary labels (i.e. 0 or 1).
The likelihoods of x are:

$$p(x|y = 0) \sim N(0.5, 2), \quad p(x|y = 1) \sim N(2, 2)$$

- a. Using maximum likelihood, what is the threshold above which you will classify a sample x with label $y = 1$?

- b. (*) Repeat section a with the following likelihoods

$$p(x|y = 0) \sim N(1.5, 1), \quad p(x|y = 1) \sim N(3, 2)$$

For which values of x will you classify its label as 1?

- c. Now we are given the priors $p(y = 0) = p(y = 1) = 0.5$

Using bayes classifier and the likelihoods from a. for which values of x will you classify its label as 1?

- d. (*) Repeat section c with the following priors:

$$p(y = 0) = 0.6, \quad p(y = 1) = 0.4$$

(*) in sections b. and d. you may use an online calculator such as wolfram alpha to get the final result.

2. As part of your job as an analyst in an investment house, you need to predict whether a certain company will distribute dividend this year, based on:

X - the company's revenue growth (in percentages) from last year.

After examining a large amount of companies, you found that the average revenue growth (X) of companies that distribute dividends is 6 percent, while the average revenue growth of companies that do not distribute dividends is 2 percent.

In addition, the variance of X for the two groups (distribute / do not distribute dividends) is 30%. It is also known that 55% of companies distribute dividends. A company named "D-Drilling" caught your attention, and on a brief examination you discovered that from last year, this company showed a 5% revenue growth. What is the probability that "D-Drilling" will distribute dividend this year? Assume that X for every group is normal and use Bayesian theory. Formulas should be displayed according to their use.

3. In lecture 10, we have an example for spam filtering using bag of words.

Denote the following table:

X_1	X_2	X_3	X_4	X_5	Y
We	Nigeria	Aspirin	Money	Help	
0	0	0	1	1	0
1	0	0	1	0	1
1	1	1	0	1	1
1	1	0	0	0	0
0	0	1	0	1	0

Using The same table as in the example, what is the predicted probability of the next sentence of being spam?

"We, At Cameroon enterprise, provide medicine (including aspirin) for those in need. Help save lives!"

Show the steps of your calculation.

Good Luck!