UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

FYS-2021 - Exercise Set 3

Bayesian Decision Theory and Parametric Methods

Department of Physics and Technology
Faculty of Science and Technology

From book (fourth edition in parenthesis)

3.1, 4.4 (4.6)

Bayesian decision theory

Problem 1

(1a) In the univariate two-class case, explain why the Bayes' classifier minimizes the probability of misclassification when the true probability density functions are known.

Problem 2

(2a) Implement a univariate Bayesian classifier for discriminating between two normal distributions with means 0 and 1, and standard-deviations 0.5 and 0.5. Assume equal prior probabilities.

(2b) Generate 1000 samples from each of the distributions, and classify these using your implementation. Calculate the confusion matrix, and plot the histograms of the samples. Illustrate the decision boundary together with the histograms. Are all points correctly classified? Why/why not?

(2c) Generate 1000 samples from a $\mathrm{Unif}(-0.87, 0.87)$ distribution, and 1000 samples from a $\mathrm{Unif}(0.13, 1.87)$. Classify these samples using the same classifier as above. Calculate the accuracy, and compare it with the previous result. You should see a drop in accuracy, even though the mean and standard-deviation of the sampling distributions are equal. Why is this the case?

Parametric methods

Problem 3

Assume we have an independent identically distributed (iid). Sample $X_1, X_2, \dots, X_n$ from the exponential density. The probability density function is given by

$$p(x) = \frac{1}{\beta} \mathrm{e}^{-x/\beta}, \ x > 0, \ \beta > 0$$

(3a) Compute $\mathrm{E}(X)$

(3b) Show that the Maximum Likelihood estimator for β is

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(3c) Find the expectation and variance of $\hat{\beta}$. Is the estimator unbiased?

An estimator is said to be *consistent*[1] if it satisfies the following:

1. The estimator is unbiased when $n \to \infty$.
2. The variance of the estimator approaches 0 when $n \to \infty$.

(3d) Is $\hat{\beta}$ a consistent estimator?

Problem 4

In this exercise we are going to use the bayesian classifier to classify whether a person has a low or high education based on their salary. Use the file `education-salary.csv`. The file contains two columns. The first has number 0,1 and 2 which corresponds to education levels: 0: elementary and junior high school (grunnskole), 1: High school (videregående) and 2: phD. The second column is the average monthly salary for an unspecified industry.

(4a) Make box plots and histograms for the three education levels.

Assume that the income at elementary school level (0), X_0, are normally distributed:

$$X_0 \sim \mathcal{N}(\mu_0, \sigma_0)$$

(4b) Derive the expression for the log likelihood function $\mathcal{L}(\mu_0 | \boldsymbol{x})$

(4c) Derive the MLE estimate for μ_0 (assuming known $\sigma = \sigma_0$).

[1] The definition of consistency from probability theory is somewhat different, and requires that the estimator converges in probability to the true parameter. However, it can be shown that the two conditions above are sufficient for consistency.

(4d) Derive an expression for the maximum a posteriori estimate, (μ_{MAP}), for the mean in monthly salaries for the elementary (0th) level. Calculate the numerical value based on the data.

(4e) Assuming that the incomes are normal distributed, implement a function that classifies which education level a person has based on a salary. You can use the MLE estimators as parameters in the normal distribution.

(4f) Predict, using bayesian classification, the education level someone would have if they had a monthly salary of 5000,-. Comment on the result.