Linnæus University
Institute of Computer Science
Prof Welf Loöwe

Examination in Machine Learning, 4DV660, 5hp

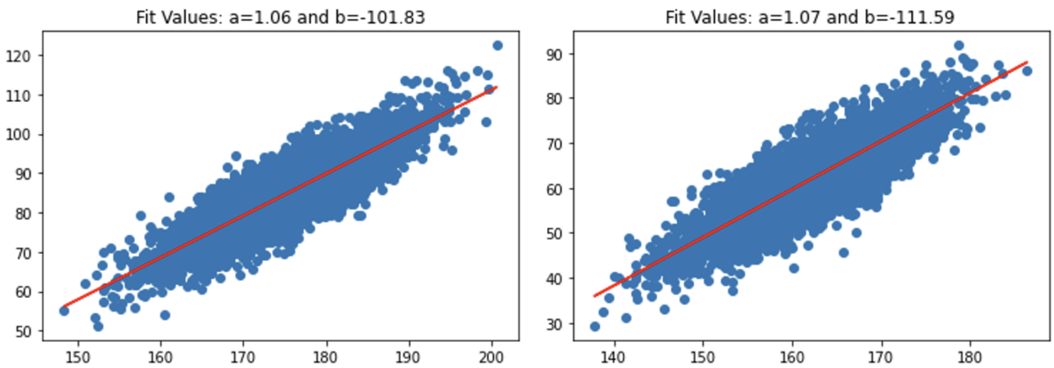
March 23, 2023, 10.00–12.00

Maximum points: 30 p
Pass: 10 p
Allowed aids: Calculator

1. Linear Regression

The two images below plot heights (in cm) and weights (in kg) of 5000 men (left) and 5000 women (right). The red lines show the respective fitted linear regressing model with their parameters on top of each image.

Figur 1: Regression plots

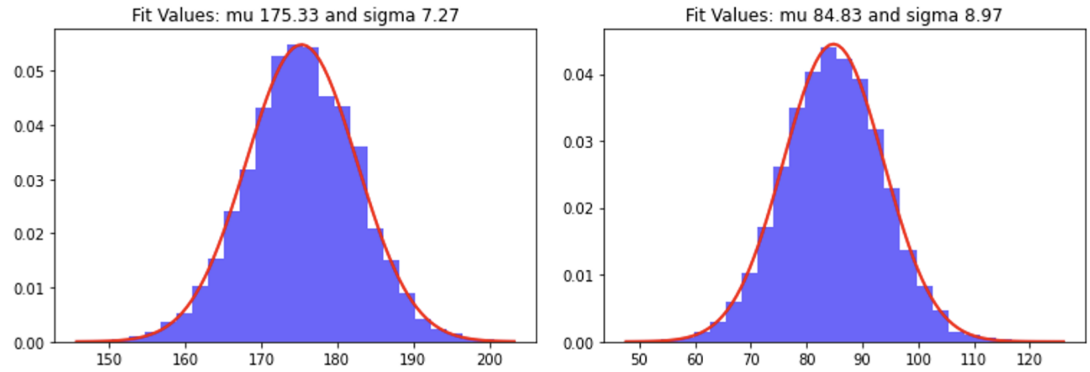


(a) What can be said about the differences of men and women based on the regression models? (5p)

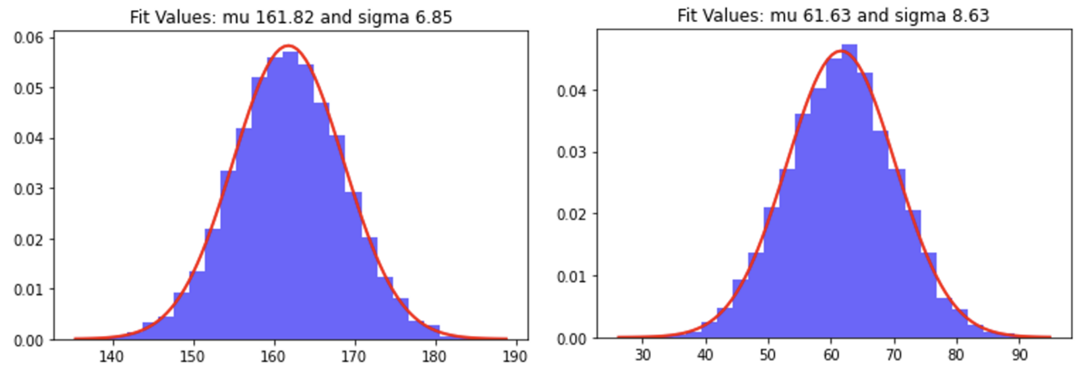
2. Naïve Bayesian Classification

Same dataset as in task 1. The two images below show the distribution of heights and weights, respectively, for men with their mean and standard deviation on top of each image.

Figur 2: Distribution plots (first for men, and second below for women



Below the same for the women:



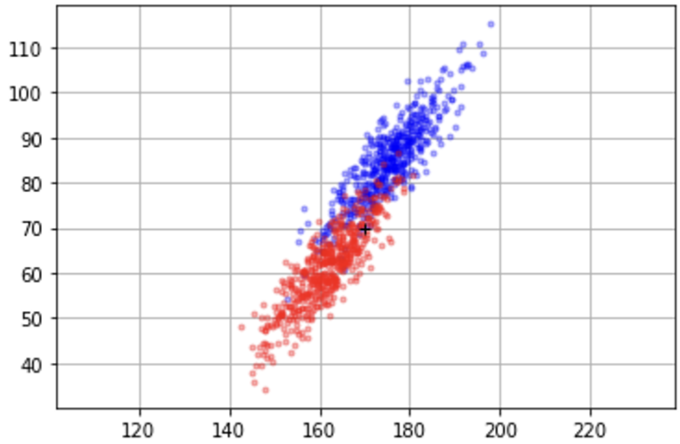
Assume you have learned a Naïve Bayesian Classifier for predicting the gender (male or female) from the training data above.

(a) Take your own weight and height: what would be the gender (male or female) that this classifier predicted? How would the classifier do that? (5p)

3. Naïve Bayesian Classification, LDA, QDA

Same dataset as in 1) and 2). The image below shows the multivariate distribution of men?s heights and weights (blue dots) and women?s heights and weights (red dots).

Figur 3: Multivariate distributionn



(a) Would you expect an improvement in the classification accuracy when using LDA instead of Naïve Bayesian Classification? Would you expect an improvement in the classification accuracy when using QDA instead of LDA? Motivate your answer. (5p)

4. Resampling and trees

Assume you have trained a regression model consisting of an ensemble of B trees using bagging based on a dataset of N datapoints. Write a short program (in your favorite programming language or in pseudocode) that estimates the test mean squared error. (5p)

You may use the following data structures and functions:

- y is an array of N response values (labels) with y[i] response for the i-th datapoint.
- mse(y1, y2) returns the mean squared error of values in the arrays y1 and y2.
- predict(i,j) returns the prediction result for the i-th datapoint using the j-th tree.
- oob(i,j) returns true if the i-th datapoint is out of bag for the j-th tree and false, otherwise.

5. Boosting

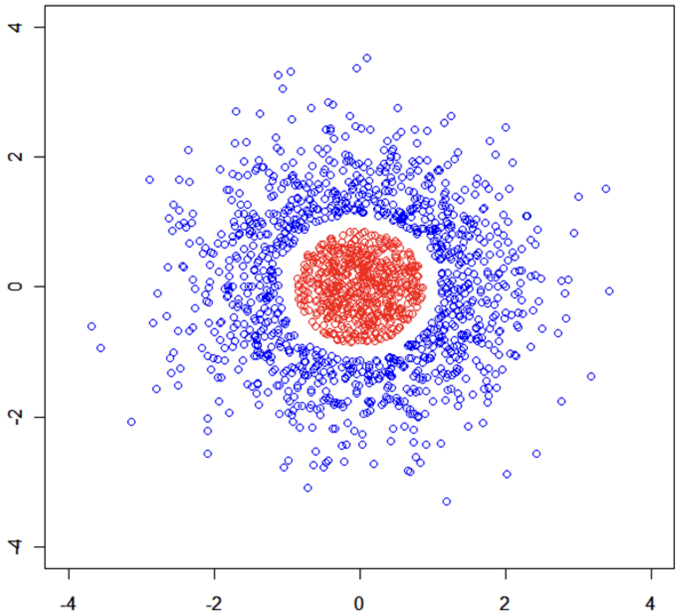
Here the boosting algorithm (as introduced on slide 40 in Lecture 7):

1. Set $\hat{f}(x) = 0$, and residual $r_i = y_i$ for all i in the training set.
2. For b=1, 2, …, B, repeat:
   a. Fit a tree $\hat{f}^b(x)$ with d splits (d+1 terminal nodes) to the training data (X, r)
   b. Update $\hat{f}$ by adding in a shrunken version (by parameter λ) of the new tree:
   $$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$
   c. Update the residuals:
   $$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$$
3. Output the boosted model:
$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x)$$

What are the hyperparameters of this algorithm? What are (reasonable) upper and lower bounds for these hyperparameters? Motivate your answer. (5p)

6. Support Vector Machines

Given these artificial datapoints (as introduced on slide 20 in Lecture 8) showing two predictor variables X1, X2 for sample datapoints from two classes red and blue.



What would be an appropriate Kernel function for an SVM classifier and why? (5p)