Examination

Linköping University, Department of Computer and Information Science, Statistics

Course code and name	TDDE01 Machine Learning
Date and time	2020-01-15, 14.00-19.00
Assisting teacher	Oleg Sysoev
Allowed aids	See “732A99_TDDE01_exam_regulations.PDF”
Grades:	5=18-20 points plus passed oral defense
	U=18-20 points plus failed oral defense
	4= 18-20 points without oral defense
	4=14-17 points with or without oral defense
	3=10-13 points with or without oral defense
	U=0-9 points with or without oral defense

Provide a detailed report that includes plots, conclusions and interpretations. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in the appendix.

Note: seed 12345 should be used in all codes that assumes randomness unless stated otherwise!

Assignment 1 (10p)

The data file **default.csv** contains information about the credit card payments of clients in the bank as well as information about whether the customer is reliable (default_payment=0, i.e. No) or unreliable (default_payment=1, i.e. Yes). The following variables are also available for the analysis:

- Limit_bal: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
- Sex (0 = male; 1 = female).
- Education (0 = graduate school; 1 = university; 2 = high school; 3 = others).
- Marriage (0 = married; 1 = single; 2 = others).
- Age (year).

1. Write a function myBayes that for a given number *k*
    a. Transforms continuous variables (Age, Limit_bal) into categorical ones with *k* categories and using function cut_interval() from library **ggplot2**
    b. Divides the data into training, validation and test (40/30/30) with the seed 12345 by using the standard codes provided in the lecture slides
    c. Learns a Naïve Bayes classifier (from package e1071) for default_payment as target, and all other variables (where Age and Limit_bal are discretized) from the training data as features, and evaluates training, validation and test misclassification errors
    d. Returns the estimated Naïve Bayes object and the estimated errors

    Use myBayes to present dependence of the training, validation and test errors on the value of k=2,3,…,10 in one plot. Find the optimal k value and interpret the plot in terms of complexity and bias-variance tradeoff. Finally, use the output of the Naïve Bayes model for k=2 to understand which feature is the most influential for prediction of the first observation in the training data. **(4p)**
2. Divide the original data into training, validation and test (40/30/30) by using the standard codes provided in the lecture slides. Implement a function that uses training data, depends on parameter vector **w** and computes the minus log-likelihood of the logistic regression model in which default_payment is target variable, Age/100 and Gender are the features. Compare the minus log-likelihood values for $\boldsymbol{w} = (w_0, w_{age}, w_{sex})$ equal to
    a. (0,1,0)
    b. (0,0,1)
    c. (1,1,1)

    What can be stated by comparing these minus log-likelihood values? **(3p)**
3. Use function from step 2 and function optim() with initial values (1,1,1) to find the optimal values of **w**. Report equation of the decision boundary corresponding to the optimal model and the test misclassification error. Compare the training and test errors and make necessary conclusions **(3p)**
    a. **Remark:** if you don't manage to solve step 2, you may use function glm() to derive the optimal model here, but the points will be reduced

Assignment 2 (10p)

KERNEL METHODS – 5 POINTS

In the slides 11 and 12 of the lecture on kernel methods, you can see how to produce a probabilistic classifier by using kernel density estimation and Bayes theorem. You are asked to implement such a classifier. First, you have to produce the learning data by running the code below, which samples 1500 points from class 1 and 1000 points from class 2. These points are stored in the variables data_class1 and data_class2. Second, you have to use the Gaussian kernel and, thus, you have to select an appropriate kernel width h. Choose a value that you deem appropriate. Choose the value manually and disregard overfitting issues. Explain your choice. Finally, compute the posterior distribution of class 1 for the points

in the interval [-5, 25], e.g. for the points in seq(-5, 25, 0.1). Visualize this distribution with a plot like the second plot in slide 12.

```
N_class1 <- 1500

N_class2 <- 1000

data_class1 <- NULL

for(i in 1:N_class1){

  a <- rbinom(n = 1, size = 1, prob = 0.3)

  b <- rnorm(n = 1, mean = 15, sd = 3) * a + (1-a) * rnorm(n = 1, mean = 4, sd = 2)

  data_class1 <- c(data_class1,b)

}

data_class2 <- NULL

for(i in 1:N_class2){

  a <- rbinom(n = 1, size = 1, prob = 0.4)

  b <- rnorm(n = 1, mean = 10, sd = 5) * a + (1-a) * rnorm(n = 1, mean = 15, sd = 2)

  data_class2 <- c(data_class2,b)

}
```

NEURAL NETWORKS – 5 POINTS

You are asked to perform model selection on the iris dataset with the package neuralnet, i.e. the same package that you used in the lab. The iris dataset is included with R. You can find information about the dataset in the help file for the function neuralnet or by typing "?iris". The dataset has 150 cases. Use 50 cases for training, 50 for validation and 50 for test. Be careful when you create these sets as the cases are sorted by class in the dataset. Use the default parameters when learning the neural networks. Select among the following five models: Using only sepal length to predict the species, or using only sepal width, or using only petal length, or using only petal width, or using only sepal length and sepal width. Estimate the generalization error of the model selected. Produce the model to return to the user. Finally answer the following question. The iris dataset is a classification problem with three classes and, thus, the neural network has to return three probabilities (one for each class) that sum up to 1. You can see this by plotting the neural network learned. You cannot do this with the sigmoid activation function. What activation function do we typically use when we have more than two classes ? The answer is not in the slides but in Bishop's book.