Examination

Linköping University, Department of Computer and Information Science, Statistics

Course code and name	732A99 Machine Learning
Date and time	2020-03-19, 8.00-13.00
Assisting teacher	Oleg Sysoev
Allowed aids	“Pattern recognition and Machine Learning” by Bishop and “The Elements of Statistical learning” by Hastie
Grades:	A=19-20 points B=16-18 points C=11-15 points D=9-10 points E=7-8 points F=0-6 points

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.

- FIRST READ THE FILE **732A99_TDDE01_EXAM_ REGULATIONS.PDF** UNLESS YOU HAVE ALREADY DONE THAT IN ADVANCE
- **Use seed 12345 when randomness is present unless specified otherwise.**
- **Specify RNGversion("3.5.2") in the code**

Assignment 1 (4p)

The data file **Dailytemperature.csv** contains information about the daily temperatures in Stockholm for some time period

1. Create new features ϕ_{1i} and ϕ_{2i} by using variable Day denoted here as x as follows:

$$\phi_{1i} = \sin(0.5^i x),\, i = -50, -49, \ldots -1, 0, 1, \ldots 50$$
$$\phi_{2i} = \cos(0.5^i x),\, i = -50, -49, \ldots -1, 0, 1, \ldots 50$$

Use all these 202 features to fit a Lasso regression with target Temperature and plot a dependence of the degrees of freedom on the value of the penalty factor. Explain the trend you observe in this dependence. By using cross-validation, present the dependence of the predicted MSE on the log-penalty factor and state whether the optimal penalty factor is statistically significantly better than log-penalty factor equal to -4. Present also the number of non-zero features corresponding to the optimal penalty factor. Finally, present the time series plot of the original and the fitted data corresponding to the optimal penalty factor and comment on the quality of fit.

Assignment 2 (6p)

In this assignment, you will work with dataset “mtcars” present in R base library. Type *mtcars* in R console to inspect the data or *?mtcars* to see the meaning of different variables.

1. Use a reduced data set with only variables *mpg* and *hp* in order to extract the direction of the first principle component: do this by using function *eigen()* and other basic R functions and report the components of the first principle component. Finally, plot the reduced data and the line that shows the first principle component direction of the data. State whether this direction looks reasonable according to what PCA is supposed to do. **(3p)**
2. Use a reduced data set with features *mpg* and *hp* and target *am* in order to perform an LDA classification. Make a scatterplot of the *mpg* versus *hp* where the data are colored by *am* and make a similar plot for the classified data. Comment on the quality of the prediction and how this is related to the assumptions of the LDA. Finally, write down a fitted probabilistic model for the LDA and use it and the package *mvtnorm* to generate a new data set with the same amount of observations per class. Plot the new data and comment whether they look like the original data and whether they should look similar. **(3p)**

Assignment 3 (10p)

ENSEMBLE METHODS – 7 POINTS

(2 p) You are asked to prove **formally** that for regression, the bagging error equals 1/B of the average of the individual errors when the individual errors have zero mean and are uncorrelated (B = number of individual regressors). **Explain** step by step your proof.

(3 p) You are now asked to show **experimentally** that, under the assumptions in the paragraph above, the bagging error is indeed smaller than the average of individual errors. To do so, assume that there are 10 individual regressors. Their individual errors are jointly distributed according to a 10-dimensional normal distribution. Each element of the diagonal of the covariance matrix is uniformly distributed in the interval [1,2]. The mean vector and the off-diagonal elements of the covariance matrix are up to you, but you

must respect the assumptions in the previous paragraph. Your task consists in generating 100 samples of the 10 individual errors, computing the bagging error and the average individual error, and plotting an histogram that shows that the bagging error is smaller in general.

(2 p) You are asked to show **experimentally** that the variance of the bagging error is smaller than the variance of the average individual error. To do so, you need to repeat the experiment above 100 times and plot the resulting variances. Finally, **answer** the following question: Why is it important that the variance of the error is as small as possible ?

NEURAL NETWORKS – 3 POINTS

The code below trains a neural network to predict the sine function in the interval [0, 10]. In the code, there are two variables: Var and Sin. The code trains a NN to predict Sin from Var. You are now asked to try to predict Var from Sin. You **must** use the same settings and training data as in the code below, i.e. you are only allowed to swap the roles of Var and Sin. **Answer** the following question: Why do you think that the new results are better/worse than the original ones ?

```
library(neuralnet)

set.seed(1234567890)

Var <- runif(50, 0, 10)

tr <- data.frame(Var, Sin=sin(Var))

winit <- runif(31, -1, 1)

nn <- neuralnet(formula = Sin ~ Var, data = tr, hidden = 10, startweights = winit, threshold = 0.02, lifesign = "full")

plot(tr[,1],predict(nn,tr), col="blue", cex=3)

points(tr, col = "red", cex=3)
```