Proef/oefen tentamen Oktober 2018, vragen

Machine Learning (Erasmus Universiteit Rotterdam)

Questions

1	2	3	4	5	6	7	8	9	10

Student ID

1	☐	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐
0	☐	☐	☐	☐	☐	☐

Name student

_____

Machine Learning (FEM31002)

Sample Exam

19 October 2018 09:00 - 12:00

Question 1

Given a data set, write down a pseudocode (the steps) that you will use to apply K-NN on a computer.

5p **1** Pseudocode.

Question 2
Derive the following equation.

10p **2** $E\left(\left(y_0 - \hat{f}(x_0)\right)^2\right) = \text{Var}\left(\hat{f}(x_0)\right) + \left(\text{Bias}\left(\hat{f}(x_0)\right)\right)^2 + \text{Var}(\epsilon)$

Question 3

Consider the manufacturing cycle of a mobile phone resulting with a production batch of $1,000$ phones. Each process is followed by an inspection step. The main goal is to detect the defective phones correctly so that the customer satisfaction does not decrease. However, the process should not label either the non-defective ones as defective or the defective ones as non-defective. To detect the defective ones, a classifier (*e.g.* LDA) is trained with a certain probability threshold τ. According to the results, at the end of a cycle, the classifier labeled 120 phones as defective whereas 10% of these phones are misclassified. The rate is 5% for the non-defective 880 phones.

2.5p **3a** Construct a confusion matrix.

		Actual	
		True	False
Predicted	True	....	....
	False	....	....

2.5p **3b** Supposed that the manufacturer has decided to pay high attention to the customer satisfaction. How should τ be changed to modify inspection process?

5p **3c** How would the change in the τ result in the confusion matrix? In which cells you are expecting a decrease? In which cells you are expecting an increase? Indicate in the table below. Mark each cell with D (decrease) or I (Increase).

		Actual	
		True	False
Predicted	True	....	....
	False	....	....

Question 4

Derive the ridge regularization from the Bayesian point of view as we have discussed through one of the assignments.

10p **4** Derivation.

Question 5
Recall that

$$\text{AIC} = \frac{1}{n\hat{\sigma}^2}\left(\text{RSS} + 2d\hat{\sigma}^2\right),$$
$$\text{BIC} = \frac{1}{n}\left(\text{RSS} + \log(n)d\hat{\sigma}^2\right).$$

10p **5** Compare AIC and BIC scores and specify when to use each one of them.

Question 6
Answer the following questions.

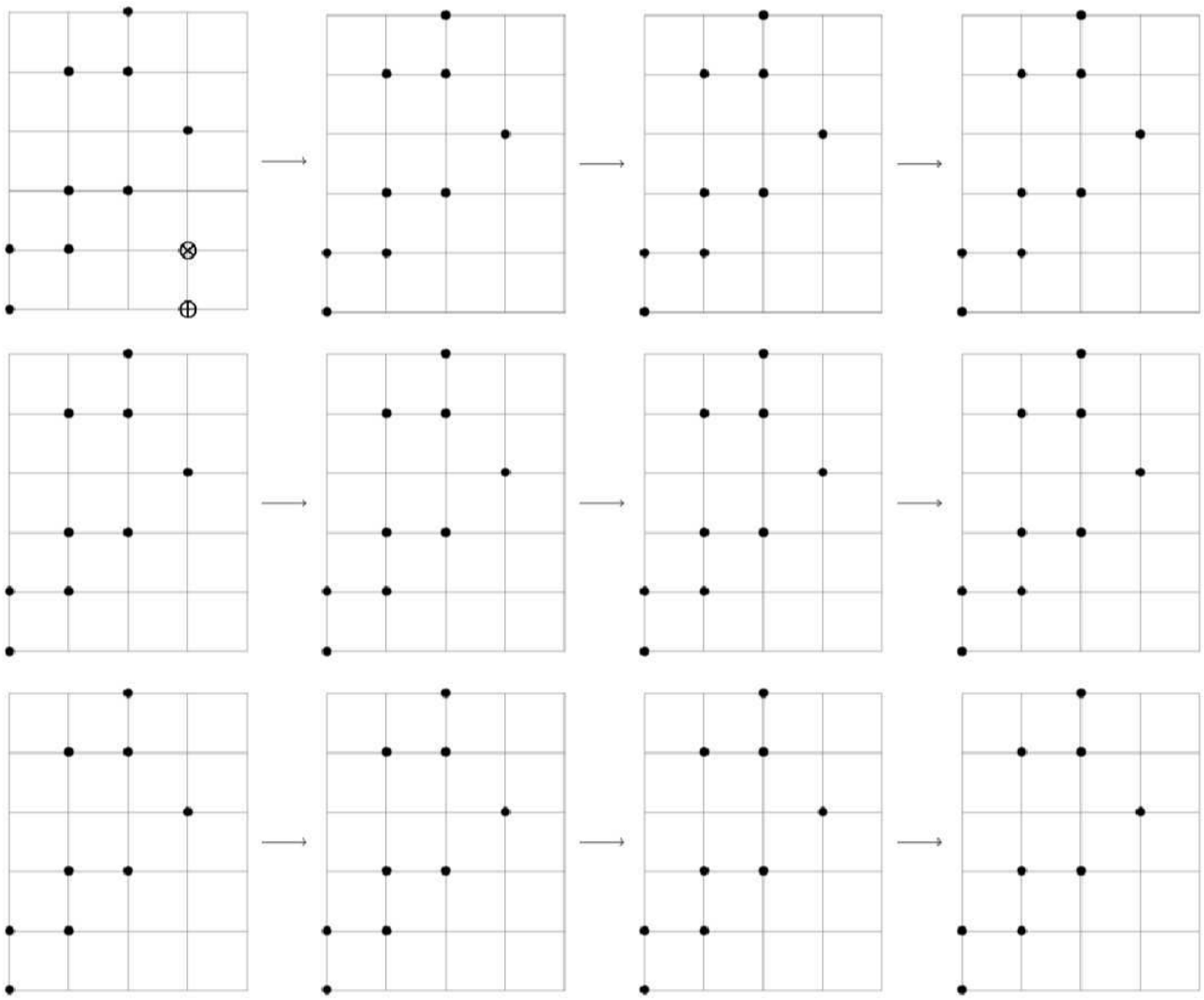
2.5p **6a** What are the two advantages of using AdaBoost as discussed in the paper entitled "*Aggregate features and AdaBoost for music classification*?"

7.5p **6b** Compare **k-fold** cross validation approach against the **leave-one-out** cross validation approach in terms of bias, variance and computation time.

Question 7

Group the points given in the grids below by using the **K-means** clustering method step by step. You are supposed to have **two** clusters in the end. Describe the members of the clusters with $(+, \times)$, and the centroids of the clusters with $(\oplus, \otimes)$. Assume that the random initial centers of the clusters are given as $\oplus = (3, 1)$ and $\otimes = (3, 0)$.

5p **7** Iterations



Question 8

Consider the likelihood function $\ell(\beta_0, \ldots, \beta_p)$ for logistic regression as we have discussed in the class.

4p **8a** Show that

$$\ell(\beta_0, \ldots, \beta_p) = \Pi_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}.$$

6p **8b** Using part a. above show that

$$\frac{\partial \log(\ell(\beta_0, \ldots, \beta_p))}{\partial \beta_j} = \sum_{i=1}^{n} (y_i - p(x_i)) x_{ij}.$$

Question 9

Suppose that you are given the data set in Table 1. Then answer the following classification tree questions.

	x_1	x_2	class
S_1	black	0	1
S_2	black	1	1
S_3	red	0	2
S_4	black	1	1
S_5	red	0	1
S_6	red	1	2
S_7	red	0	2
S_8	black	1	1
S_9	red	1	1
S_{10}	red	0	1

Table 1: Dataset

2p **9a** Specify the splits in your tree.

4p **9b** Calculate explicitly the gain for each split according to **Gini** index. Which split should be selected?

4p **9c** Calculate explicitly the gain for each split according to **Entropy**. Which split should be selected?

Question 10

For each of parts a. through d. below, indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.

5p **10a** The sample size n is extremely large, and the number of predictors p is small.

5p **10b** The number of predictors p is extremely large, and the number of observations n is small.

5p **10c** The relationship between the predictors and response is highly nonlinear.

5p **10d** The variance of the error terms, *i.e.* Var(ϵ), is extremely high.
