EURECOM
Machine Learning and Intelligent Systems

Final exam

Fall 2021

April 2022

This exam contains 6 pages (including this cover page) and 6 questions. Question 6 is optional and acts as a bonus.

Total of points is 20.

Distribution of Marks

Question	Points	Score
1	5	
2	5	
3	2½	
4	3½	
5	4	
6	0	
Total:	20	

1. **Global overview**
The following questions require a short answer (2-3 sentences). To discourage random guessing, no credit is given to answers without a correct explanation.

(a) [½ point] In one sentence, tell the difference between classification and regression.

(b) [½ point] Consider the following logistic regression classifier of the form

$$\mathcal{P}(y = 1|\mathbf{x}, \mathbf{w}) = \sigma(w_0 + \mathbf{w}^T\mathbf{x}),$$

where $\sigma(\cdot)$ is the sigmoid function and $\mathbf{x} \in \mathbb{R}^2$. $\mathcal{P}(y = 1|\mathbf{x}, \mathbf{w})$ can be viewed as a function of $\mathbf{x}$, that we can get by changing the parameters $\mathbf{w}$. What would be the range of $\mathcal{P}$ in such case?

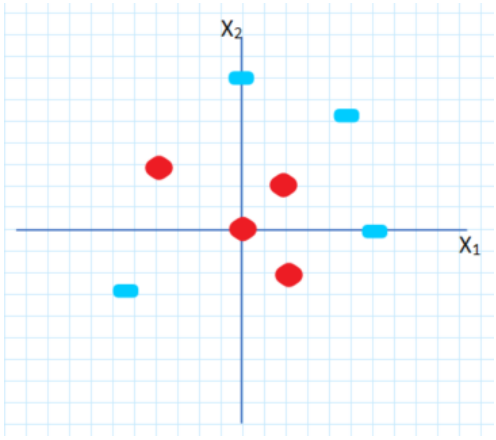
(c) [½ point] **True/False.** Given $\mathbf{w}$ the weights of the perceptron, we can consider that a data point $\mathbf{x}$ is correctly classified if and only if $\mathbf{w}^T\mathbf{x}y > 0$.

(d) [½ point] **Yes/No.**You train an SVM using N training points. Your observe that the trained model has M support vectors. A new set of K points arrives. You retrain your SVM using $N + K$ points. Can you tell how many support vectors your new model will have?

(e) [½ point] **True/False.** In ridge regression, choosing a very high value for the regularization term (λ) will lead to a model with high bias and high variance.

(f) [½ point] What is the role of the validation and testing data sets?

(g) [1 point] You want to classify the data set below with a linear SVM. You use a linear, a quadratic, a cubic and a Gaussian (RBF) kernel. List the kernels that can perfectly classify the data.



(h) [½ point] **True/False.** A weak classifier will never have an error rate greater than 1/2.

(i) [½ point] We are training a decision tree using a data set consisting of two features x_1, x_2 and their label y as illustrated below. Which feature leads to the best split?

x_1	x_2	y
0	1	+1
1	0	-1
0	1	+1
1	1	+1

2. **Linear regression**

You have a data set with N samples $\{(x_i, y_i)\}_{i=1}^N$, $x_i \in \mathbb{R}, y_i \in \mathbb{R}$

(a) [2 points] You use linear regression to fit the data. To test the resulting model, you split your data into training and testing. You register the mean training and testing errors of this first model.

Now, you start to increase your training set gradually. As the training set size increases, what do you expect will happen with the mean training error? Will it increase or decrease? and the mean testing error? Explain your answer.

(b) [3 points] You decide to change the assumption about the model that fits your data. The model has one unknown parameter w to be learned from data:

$$y_i \sim \mathcal{N}(\log(wx_i), 1)$$

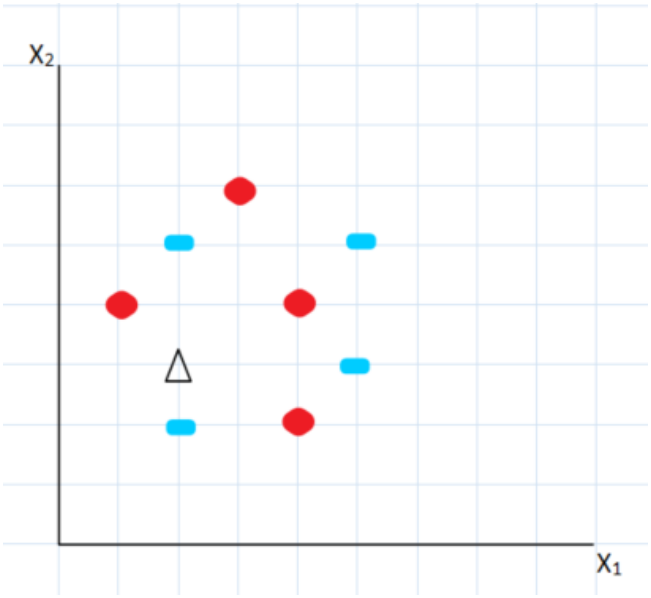
Note that the variance is known and equal to one. You use maximum likelihood estimation to obtain w. You do the math and you come to the conclusion that w has to satisfy one of the following equations. Which one? No credits will be given without the derivation.

1. $\sum_i^N y_i = \sum_i^N \log(wx_i)$
2. $w = \sum_i^N \log(y_i x_i)$
3. $\sum_i^N y_i x_i = \sum_i^N y_i x_i \log(wx_i)$
4. $\sum_i^N x_i = \sum_i^N y_i \log(wx_i)$

Hint: Follow the MLE HOWTO

3. **KNNs**

Consider the training set with two features X_1, X_2 and a label y denoting two clases, circles and bars, as in the plot below. You want to classify the test point (triangle) using KNNs.



(a) [2 points] Using the Euclidean distance as metric, how is the test point classified according to K? You may assume that the grid lines represent a unit. Fill in your answer in the table below:

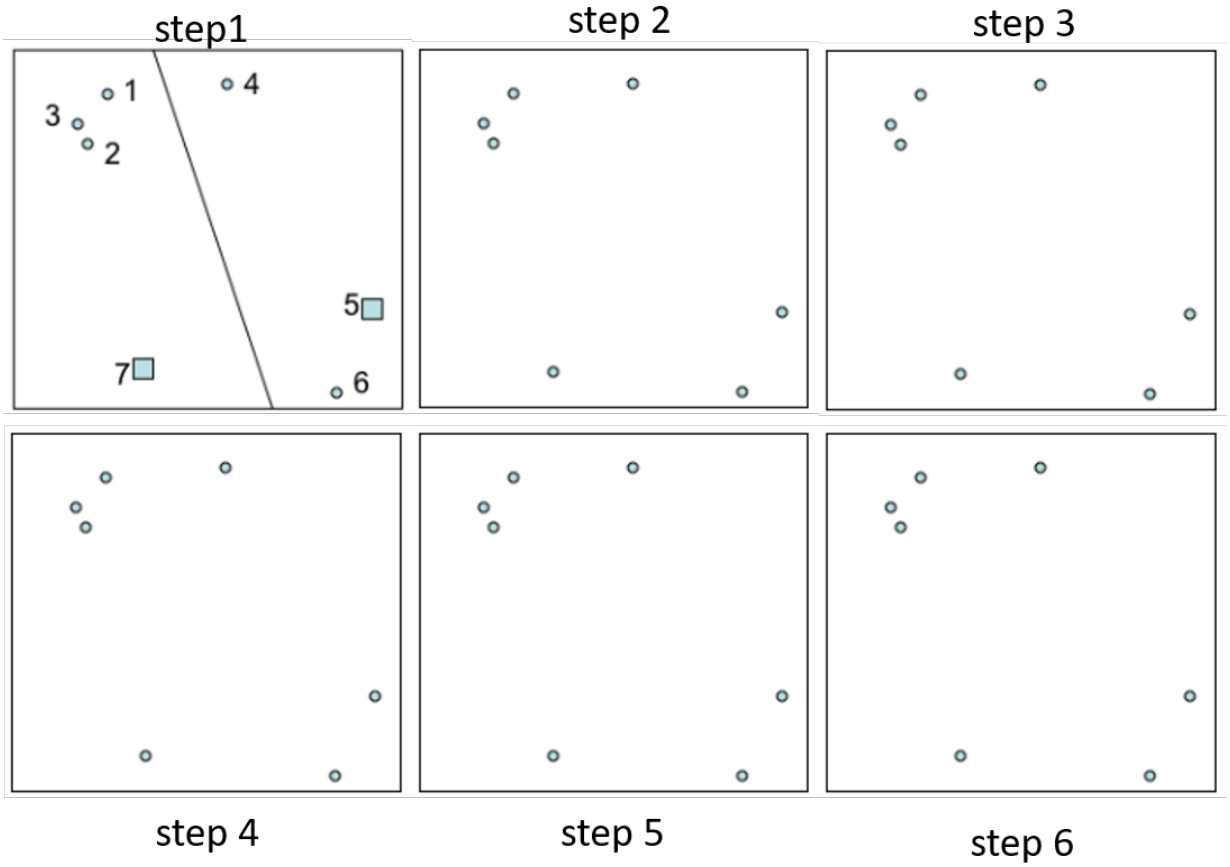
K	$\hat{y}$ (output)
1	
2	
3	
4	
6	
8	

(b) [½ point] For which values of K will KNN be more prone to have high bias? high variance?

4. **K-Means**

(a) [½ point] Give one advantage of hierarchical clustering over K-means and one advantage of K-means over hierarchical clustering.

(b) [3 points] Consider the data set below, consisting of unlabeled circles. You will perform K-means on it. The algorithm has already been initialized (step 1) by randomly picking two points of the dataset (points 5 and 7) as the two initial centroids. At every step, draw the centroids as squares and the decision boundaries that define each cluster. If no points belong to a particular cluster, assume its center does not change. Use as many of the pictures as you need for convergence. You should not need more squares than the ones provided



5. **Anomaly detection**

You are the new hire of a company facing problems with intrusion attacks to their network. They believe that, as a machine learning expert, you can help them automate the anomaly (i.e. attack) detection process, which is currently done manually.
You consider that you can use the concepts learned in MALIS to solve the problem

(a) [1 point] Would you use a supervised or an unsupervised approach? Explain

(b) [1 point] Which method would you use? Which advantages it has? Disadvantages?

(c) [2 points] Describe the algorithm you would put in place. Be very specific about all the steps