Machine Learning & Machine Learning (extended)

(Practice Exercise Sheet)

Please put in some effort in to attempting to solve these exercises to get the most out of topics studied and to increase your chance of doing well on the Canvas Quiz test. In addition, you should attempt the exercise questions specified towards the end of each lecture slides.

This document is only an example of the kind of questions you may be asked in the Quiz, but it's no way an exhaustive list of all kind of questions.

Some of these questions will be solved in the class on Thursday as preparation before the Quiz.

Linear Regression/Modelling

Question: Suppose we have a set of values as below. Is it possible to find a function f such that y=f(x)? Can you describe this function f and suggest how it can be estimated from the data given?

x	y
1.2	1.1
2.3	2.1
3.0	3.1
3.8	4.0
4.7	4.9
5.9	5.9

Question: What is model complexity in linear modelling? If the model complexity increases, does it help or hinder function modelling? Comment on both scenarios.

Question: Given the linear models below to predict winning time t from Olympic year x for 100m race:

Men's model: t = 36.4165 – 0.0133x

Women's model: t = 40.9242 – 0.0151x

a) Can you predict what year the Women's winning time will be lower than the Men's winning time?

b) What are the predicted winning times for both Men's race and Women's race for that year? Comment on that.

Question: Let's consider that a linear model (as a line) has two parameters $\widehat{w}_0$ and $\widehat{w}_1$ which can be estimated as below:

$\widehat{w}_0 = \bar{t} - \widehat{w}_1 \bar{x}$

$\widehat{w}_1 = \frac{\overline{xt} - \bar{x}\bar{t}}{\overline{x^2} - (\bar{x})^2}$

a) Given the data below, can you estimate the model parameters using above equations.

x	t
1.2	1.1
2.3	2.1
3.0	3.1
3.8	4.0
4.7	4.9
5.9	5.9

b) Having found the model parameters, can you use the model to predict target labels t?

c) Is there a way to find how good the model is doing by comparing model predictions with actual targets?

Question: Given the total training loss estimate function below between actual target labels t_n and model predicted target labels $\boldsymbol{w}^T \boldsymbol{x}_n$ where $\boldsymbol{x}_n$ denote the attributes and $\boldsymbol{w}$ denote model parameters:

$$\mathcal{L} = \sum_{n=1}^{N} (t_n - \boldsymbol{w}^T \boldsymbol{x}_n)^2 = (\boldsymbol{t} - \boldsymbol{X}\boldsymbol{w})^T (\boldsymbol{t} - \boldsymbol{X}\boldsymbol{w})$$

Can you derive the optimal parameters estimate $\widehat{\boldsymbol{w}}$ for linear modelling that minimizes this total loss?

Question: A linear model can attempt to consider the noise present in the following way:

$t_n = \boldsymbol{w}^T \boldsymbol{x}_n + \varepsilon_n$ where ε_n denotes the additive noise.

a) Why is the noise considered to be additive?

b) Is there a way to model the statistical distribution of this noise?

Bayesian Classification

Question: A patient is tested by a lab test for a disease that has prevalence of 1 in 1000 in the population. The lab test has a false-positive rate of 1% and a false negative rate of 1%.

a) If the lab test result is positive, what is the probability that the patient actually has the disease?

b) Is it more probable that the patient has the disease or not?

c) Would the answers to a) and b) differ if a maximum likelihood versus maximum a posteriori hypothesis estimation method is used? Comment on your answer.

Question: Consider the following past examples available to train a robot cleaner to predict whether or not an office contains recycling bin.

	Office Status	Floor	Department	Office	Recycle Bin?
1	Faculty	4	CS	Medium	Y
2	Student	4	EE	Large	Y
3	Staff	5	CS	Medium	N
4	Student	3	EE	Small	Y
5	Staff	4	CS	Medium	N

How would a naïve Bayes classifier classify the following instance?

	Office Status	Floor	Department	Office	Recycle Bin?
6	Student	4	CS	Small	?

Question: Assume we have a data set described by the following three variables (i.e. attributes):

Hair = {B,D}, where B=blonde, D=dark.

Height = {T,S}, where T=tall, S=short.

Country = {G,P}, where G=Greenland, P=Poland.

You are given the following training data set (Hair, Height, Country):

(B,T,G) (B,T,G) (B,T,P)
(D,T,G) (D,T,G) (B,T,P)
(D,T,G) (D,T,G) (B,T,P)
(D,T,G) (D,T,G) (D,T,P)
(B,T,G) (B,T,G) (D,T,P)
(B,S,G) (B,S,G) (D,S,P)
(B,S,G) (B,S,G) (B,S,P)
(D,S,G) (D,S,G) (D,S,P)

We want to answer the following question: If you observe a new individual who is tall with blonde hair, what is the most likely country of origin?

a) Find the maximum a posteriori (MAP) estimate to the above question, using the Naïve Bayes assumption. Show all of your working.

b) Find the maximum a posteriori (MAP) estimate to the above question, without using the Naïve Bayes assumption. Show all of your working.

c) Find the Maximum Likelihood (ML) estimate to the above question, using the Naïve Bayes assumption.

d) Find the Maximum Likelihood (ML) estimate to the above question, without using the Naïve Bayes assumption.

e) Which of the above methods would you trust in each of the following situations, and why?

- a large number of training examples described by a small number of attributes
- a small number of training examples described by a large number of attributes

f) Explain how would you solve the same question if instead of blonde/dark we would have a continuous valued measurement of the hair colour, and instead of tall/short we would have the actual height in centimetres?

Question: We have the following past examples data available about the students' performance.

Student ID	First-class last year?	Gender	Hard-working	Drinks?	First class next year?
1	Y	M	N	Y	Y
2	Y	M	Y	N	Y
3	N	F	Y	N	Y
4	N	M	N	Y	N
5	Y	F	Y	Y	Y
6	N	M	Y	Y	N

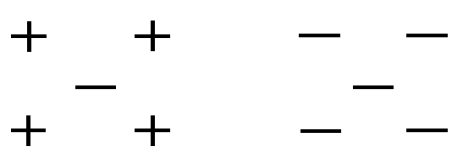
Can we predict if students 7 & 8 will earn first-class grade next year?

Student ID	First-class last year?	Gender	Hard-working	Drinks?	First class next year?
7	N	Y	N	Y	?
8	N	N	Y	Y	?

Build a Bayesian classifier to find the answer, show all the working.

k-NN Classification

Question: Consider the following data set with two real-valued input attributes x and y (i.e. the coordinates of the points) and one binary output t (taking values + or -). We want to use k-nearest neighbours (K-NN) with Euclidean distance to predict t.



a) Calculate the leave-one-out cross-validation error of 1-NN on this data set.

b) Calculate the leave-one-out cross-validation error of 3-NN on this data set.

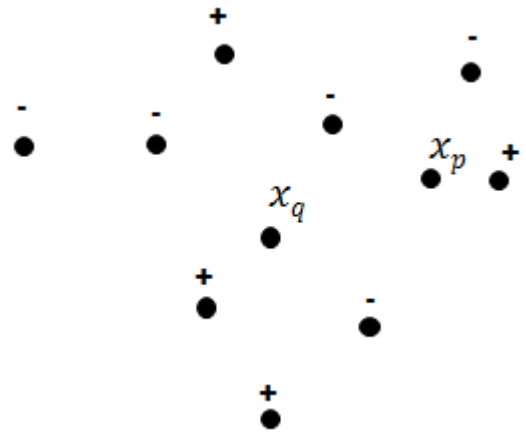
c) Describe how you would choose the number of neighbours K in K-NN in general?

Question: Which of the following increases the likelihood of over-fitting and/or under-fitting? Why?

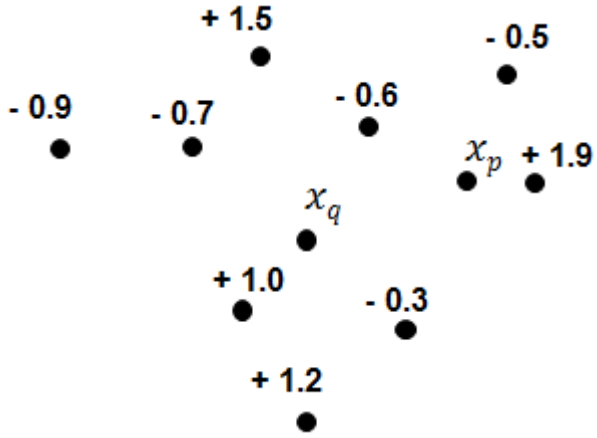
a) increasing the number of neighbours k in kNN?

b) decreasing the number of neighbours k in kNN?

Question: Assume a Boolean target function (i.e. binary classifier) and a two dimensional instance space. Determine how the kNN would classify the test instances x_p and x_q for $k = 1$, $k = 3$ and $k = 5$.



Question: In the diagram below, the numbers refer to the values taken by a real-valued target function. Calculate the values predicted for the target function at the test points x_p and x_q by kNN, with $k = 1$, $k = 3$, and $k = 5$.



Question: How do you compare weighted kNN classifier with a regular non-weighted kNN classifier? Has one of them any advantages (or disadvantages) over the other?

Evaluation Metrics for Classification

Question: What is the disadvantage of ROC curve for analysing the classification performance?

Question: Given the confusion matrix below as a measure of classifier's performance to predict class labels for a 2-class problem, answer the following:

		Actual class label	
Predicted label		Diseased	Healthy
	Diseased	21	3
	Healthy	4	37

a) Compute the overall classification accuracy?

b) Compute the classification accuracy, true positive, true negative, false positive, false negative?

c) Compute the sensitivity and specificity?

Question: Given the confusion matrix below as a measure of classifier's performance to predict class labels for a 3-class problem, answer the following:

		Actual class label		
Predicted label		Metabolic disease	Heart disease	Healthy
	Metabolic disease	34	9	1
	Heart disease	7	39	5
	Healthy	3	2	49

a) Compute the overall classification accuracy?

b) Compute the classification accuracy, true positive, true negative, false positive, false negative for each class?

c) Compute the sensitivity and specificity for each class?

Question: The ROC curve is a way to visualise the combined sensitivity and specificity performance of a classification system. Answer the following:

a) Is there a way to quantitatively measure the estimate of combined sensitivity and specificity performance from the ROC curve?

b) By analysis of ROC curve, is there a way to 'perturb' the algorithm such that the sensitivity and/or specificity can be enhanced?