NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
MACHINE LEARNING
PROGRESS EXAMINATION, DECEMBER 2020

Instructions:

- Time allocated: **90 minutes.**
- Please write your name and your serial number (if you have one).
- In what follows, Z1, Z2 and Z3 are three non-zero one-digit integers chosen at will from your University serial number. For example, if your serial number is ds2200013, you can choose Z1=1, Z2=3, Z3=2. If you don't have a serial number, chose three non-zero one-digit integers at will.
- Write down your choices for Z1, Z2, Z3 in the beginning of your paper.
- In a few of the multiple choice questions, there is more than one correct answer.
- When you finish, scan or take a photo of your paper and send it by replying to the same e-mail message by which you received the questions.

1. A potential advantage of the ridge regression method over the least squares method is:

    a. Ridge regression eliminates bias from the estimation
    [✓ b. We can avoid overfitting using a regularization term]
    c. Prior knowledge related to the specific problem is utilized
    [✓ d. Ridge regression produces a smoother fitting curve]

2. We are given a set of 200 pairs (x_i,y_i), i=1,…,200 and we seek to perform generalized linear regression using the ridge regression method. In truth, data are generated by a 6th degree polynomial in *x* with added noise. We employ 10-fold cross validation and use our well known formula for estimating the parameter vector $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \left(\Phi^{\mathrm{T}}\Phi + \lambda \mathrm{I}\right)^{-1}\Phi^{\mathrm{T}}\boldsymbol{y}$$

    In the above formula, the number of rows in matrix Φ is:

    a. 7
    b. 200
    c. 8
    [✓ d. 180]
    e. 6
    f. 190

3. In the previous problem, the number of columns in matrix Φ is:

    [✓ a. 7]
    b. 200
    c. 8
    d. 180
    e. 6
    f. 190

4. In the previous problem, the number of columns in the unit matrix I is:

    [✓ a. 7]
    b. 200
    c. 8
    d. 180
    e. 6
    f. 190

5. Consider a generalized regression problem where data points are generated by a 4th degree polynomial. Characterize as small or large the bias and variance of the estimates produced by the following models:

Model	**Bias**	**Variance**
Linear	Small/Large	Small/Large
4th degree polynomial	Small/Large	Small/Large
20th degree polynomial	Small/Large	Small/Large

6. A k-nearest neighbor classifier is used to classify instances into 2 classes. It is advisable that k is
   a. Even
   b. Prime
   c. As large as possible
   d. Odd

7. Which of the following statements is true?
   a. It is possible that a biased estimator perform better than an unbiased estimator
   b. It is not possible that a biased estimator perform better than an unbiased estimator

8. In a linear regression task, the Maximum Likelihood method can give different results from the Least Squares method when
   a. The number of dimensions exceeds the number of instances
   b. The noise is white
   c. There are outliers in our data
   d. The noise is not white
   e. There are too few instances in the training set

9. When we estimate the probability density function of a distribution based on samples drawn from the distribution, the result of the Maximum a Posteriori Probability (MAP) method tends to approach the result of the Maximum Likelihood method when:
   a. The number of patterns in the training set approaches infinity
   b. There are too few patterns in the training set
   c. Our a priori estimate for the parameter in MAP has a large standard deviation
   d. Our a priori estimate for the parameter in MAP has a small standard deviation

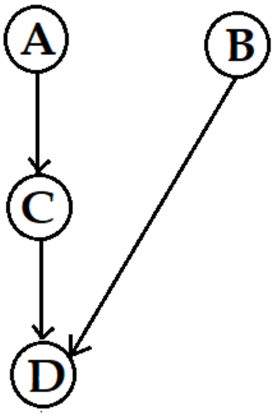
10. We wish to classify the following 2-D patterns: $\begin{bmatrix}0\\3\end{bmatrix}$ and $\begin{bmatrix}3\\4\end{bmatrix}$ to a class ω1 (target +1) and : $\begin{bmatrix}2\\1\end{bmatrix}$ to another class ω2 (target -1) using the perceptron algorithm. Starting from initial weights Z1, Z2 and bias Z3, perform **one** epoch of the perceptron algorithm (incremental, ε=1) and tabulate your results. Give the equation of the decision line that you found. Which patterns in the training set are correctly classified and which wrongly? Find the distance of pattern $\begin{bmatrix}0\\3\end{bmatrix}$ from the decision line.

11. Does a 2-neuron Hopfield network with weight matrix $\begin{bmatrix}Z1 & Z2/2\\Z2/2 & Z3\end{bmatrix}$ retrieve pattern $\begin{bmatrix}1\\-1\end{bmatrix}$ correctly?

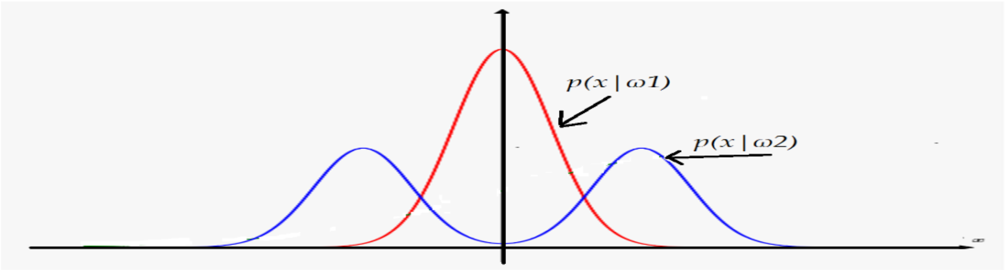
12. The two-dimensional patterns from two equiprobable classes ω_1 and ω_2 originate from gaussian distributions with means $\boldsymbol{\mu_1} = \begin{bmatrix}0\\0\end{bmatrix}$ and $\boldsymbol{\mu_2} = \begin{bmatrix}4\\4\end{bmatrix}$ respectively and common covariance matrix Σ. The eigenvectors of Σ are $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ with corresponding eigenvalues Z2 and Z3. We want to classify pattern $\boldsymbol{x} = \begin{bmatrix}4\\3\end{bmatrix}$ using a Bayes classifier. Draw a diagram to show qualitatively the isolevel curves for the two classes, that pass through point $\boldsymbol{x}$. From your diagram, decide to which class you would classify $\boldsymbol{x}$.

13. In the Bayesian network shown below, all variables are binary. For example, variable A can take the values A_1 (A is true) and A_0 (A is false).

    a. For each of the nodes A,B,C,D which are the conditional probabilities associated with the network that you should know in order to be able to perform inference?
    b. Which formula gives the joint probability P(A,B,C,D) for this network?
    c. Inference: Find $P(D_1|A_1)$ and $P(A_1|D_0)$ in terms of the conditional probabilities of question (a).



14. Probability density functions are shown for 2 equiprobable categories ω_1 and ω_2. Show which intervals of the x axis are classified by a Bayes classifier to each of the two categories:
    a. When the two categories are equally important
    b. When it is doubly damaging to misclassify patterns belonging to category ω_2.



END OF EXAM

$Z_1 = 2, \quad Z_2 = 2, \quad Z_3 = 3$

PROBLEM 10

	x_1	x_2	x_0	tx_1	tx_2	tx_0
1	0	3	-1	0	3	-1
2	3	4	-1	3	4	-1
3	2	1	-1	-2	-1	1

$$\underline{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_0 \end{pmatrix}$$

	w_1	w_2	w_0	tx_1	tx_2	tx_0	$t\underline{w}\cdot\underline{x}$	update	$\Delta\underline{w}$		
	2	2	3	0	3	-1	3	NO	0	0	0
	2	2	3	3	4	-1	11	NO	0	0	0
	2	2	3	-2	-1	1	-3	YES	-2	-1	1
Final:	0	1	4								

$\underline{w} = \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix}$ Decision line : $w_1 x_1 + w_2 x_2 - w_0 = 0$

$\Rightarrow x_2 = 4$

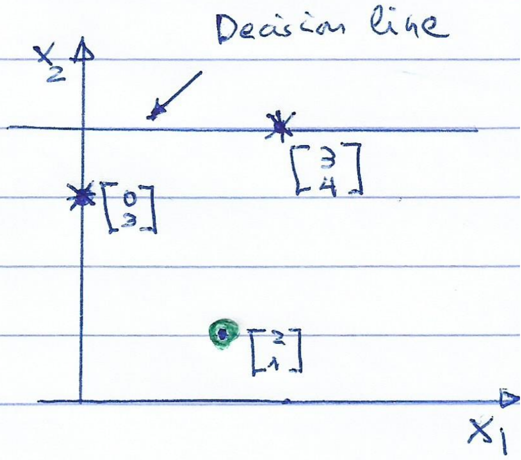
Pattern 1: $\underline{w}\cdot\underline{x} = 12 - 4 = 8 > 0$ → Correct

Pattern 2: $\underline{w}\cdot\underline{x} = 4 - 4 = 0$ → On the decision line

Pattern 3: $\underline{w}\cdot\underline{x} = 1 - 4 = -3 < 0$ ⇒ Wrong

Distance of pattern $\begin{bmatrix} 0 \\ 3 \end{bmatrix}$ from decision line:

$$D = \frac{|\underline{w}\cdot\underline{x}|}{\sqrt{w_1^2 + w_2^2}} = \frac{|3-4|}{1} = 1$$


Decision line
x_2
$\begin{bmatrix} 3 \\ 4 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 3 \end{bmatrix}$
$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$
x_1

PROBLEM 11

$W = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$, $\underline{s} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Network output $\underline{s}' = \text{sign}(W\underline{s}) = \text{cign}\left[\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right] = \text{sign}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \underline{s}$

The pattern $\underline{s}$ is retrieved correctly.

PROBLEM 12

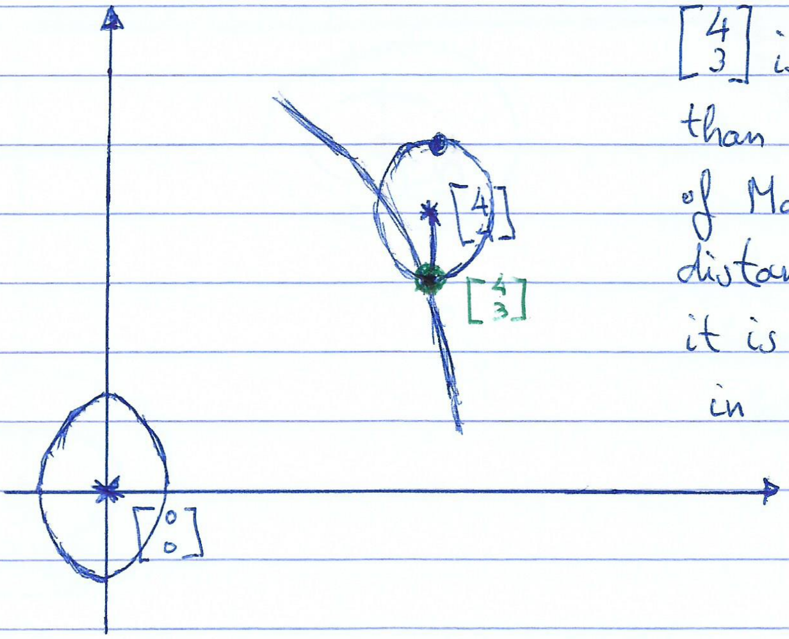
Eigenvalues of Σ: $\lambda_1 = 2$, $\lambda_2 = 3$

Corresponding eigenvectors: $\underline{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\underline{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Isolevel curves are ellipses, with axes parallel to the axes of the cartesian coordinate system.

The lengths of the axes for the ellipses are proportional to $\sqrt{2}$ and $\sqrt{3}$ respectively.

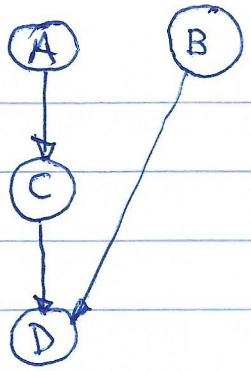
$\sqrt{3}/\sqrt{2} \simeq 1.22$



$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$
$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$ is closer to $\begin{bmatrix} 4 \\ 4 \end{bmatrix}$ than $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ in terms of Mahalanobis distance. Therefore it is classified in ω_2.

PROBLEM 13



a) $P(A_1)$

$P(B_1)$

$P(C_1|A_1)$, $P(C_0, A_1)$

$P(D_1|C_1, B_1)$, $P(D_1|C_1, B_0)$, $P(D_1|C_0, B_1)$, $P(D_1|C_0, B_0)$.

Complementaries easily computed. E.g. $P(D_0|C_1,B_0) = 1 - P(D_1|C_1, B_0)$.

b) $P(A,B,C,D) = P(D|B,C)\,P(C|A)\,P(B)\,P(A)$

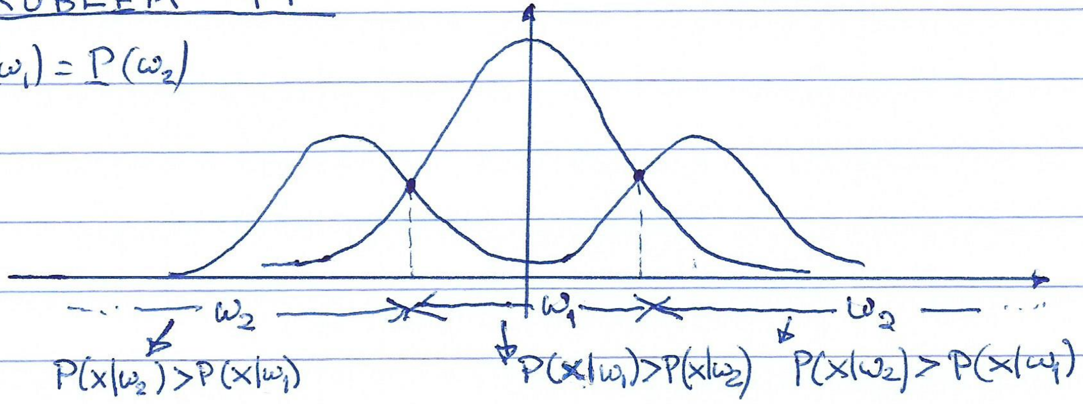
c) $$P(D_1|A_1) = \frac{P(D_1, A_1)}{P(A_1)} = \frac{\sum_{B,C} P(A_1, B, C, D_1)}{P(A_1)} \rightarrow 4 \text{ terms}$$

$$P(A_1|D_0) = \frac{P(A_1, D_0)}{P(D_0)} = \frac{\sum_{B,C} P(A_1, B, C, D_0)}{\sum_{A,B,C} P(A, B, C, D_0)} \begin{array}{l} \rightarrow 4 \text{ terms} \\ \rightarrow 8 \text{ terms.} \end{array}$$

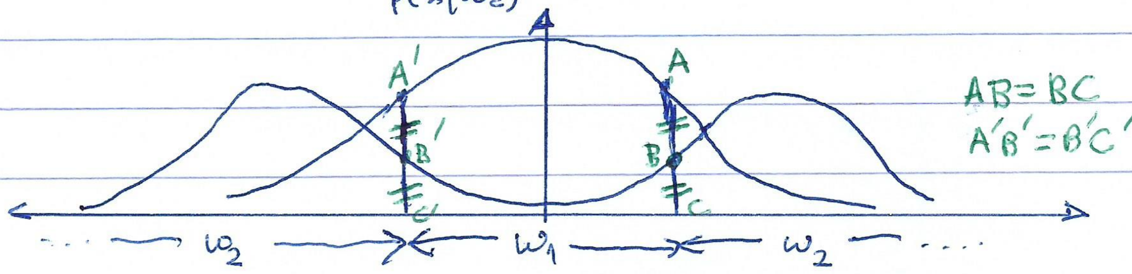
E.g. $P(D_1, A_1) = P(D_1, B_1, C_1, A_1) + P(D_1, B_1, C_0, A_1) +$
$+ P(D_1, B_0, C_1, A_1) + P(D_1, B_0, C_0, A_1)$

PROBLEM 14

a) $P(\omega_1) = P(\omega_2)$


ω_2
ω_1
ω_2
$P(x|\omega_2) > P(x|\omega_1)$
$P(x|\omega_1) > P(x|\omega_2)$
$P(x|\omega_2) > P(x|\omega_1)$

b) Classification to ω_1 : $\frac{P(x|\omega_1)}{P(x|\omega_2)} > 2$


A'
B'
C'
A
B
C
ω_2
ω_1
ω_2
$AB = BC$
$A'B' = B'C'$