NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
MACHINE LEARNING
PROGRESS EXAMINATION, DECEMBER 2024

Instructions:

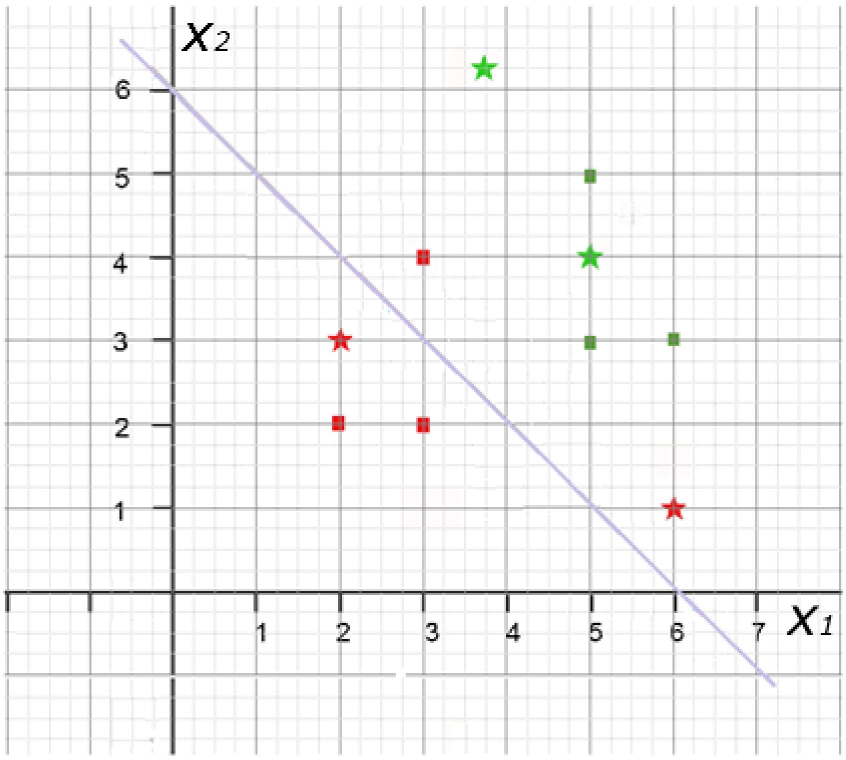
- Time allocated: **2 hours and 15 minutes.**
- Please write your name and your serial number.
- When you finish, scan your paper, group all pages into a single pdf file and upload it on the following slot in eclass: **“Assignments -> Mid Term Exam December 2024 Submissions”.**

Question 1

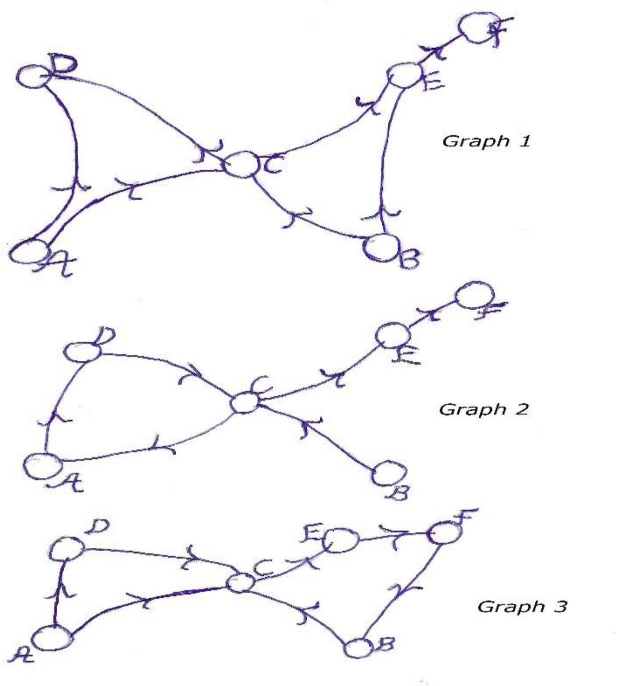
In the following diagram you can see a classification problem with two classes. Square dots belong to the training set, while stars belong to the test set. All red symbols belong in reality to class 1, while all green symbols belong to class 2.

We have run the perceptron algorithm for several epochs and have reached the separating line shown in the diagram. Throughout, we have used targets +1 for class 1 (red) and -1 for class 2 (green). At this point, there is only one out of the 6 instances in the training set which is wrongly classified.

a. Looking at the diagram and assuming that the bias w_0 is equal to 6, give the current values for the components w_1 and w_2 of the weight vector. Draw the weight vector on the diagram. *Be careful to get the signs right.*
b. Starting from the current situation, perform **one** epoch of the perceptron algorithm (incremental mode) on the training set with the learning rate parameter ε equal to 0.05. Show your effort in the usual tabular form.
c. Draw the new separating line and the new weight vector.
d. Calculate the precision and recall for the training set and for the test set before performing the extra epoch and after performing it (8 numbers in total). For the purposes of this calculation, class 1 (red) counts as the positive class. Rely on the graphs for counting true/false positives/negatives.



Question 2



In the graphs on the left, all variables are assumed to be binary. For example, variable A can take the values A_1 (A is true) and A_0 (A is false).

a. Which of the depicted three graphs represents a Bayesian network? Why?

b. For the graph which does indeed represent a Bayesian network, for each of the nodes A,B,C,D,E,F, which are the (conditional) probabilities associated with the network that you should know in order to be able to perform inference?

c. Which formula gives the joint probability $P(A,B,C,D,E,F)$ for this network?

d. Inference: Find $P(F_1|A_1)$ and $P(A_1|F_1)$ in terms of the (conditional) probabilities of question 2b.

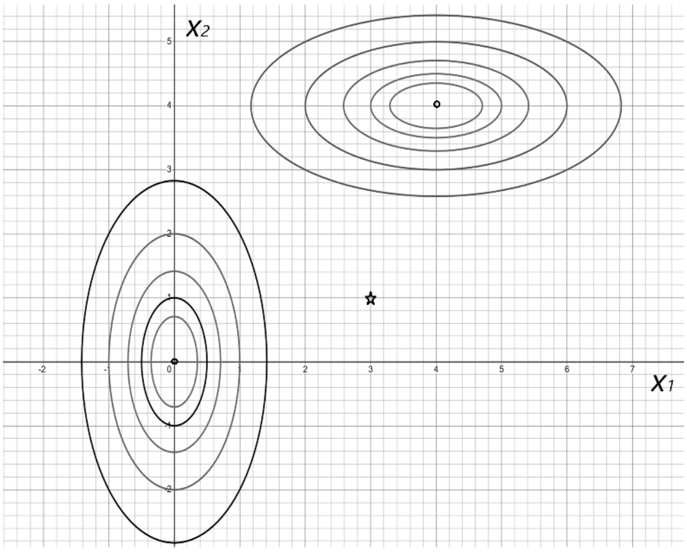
Question 3

The diagram depicts isolevel curves for the Gaussian probability density functions $p(\mathbf{x}|A)$ and $p(\mathbf{x}|B)$ of two equiprobable classes A and B in two-dimensional space. Assume that for both distributions, the smallest eigenvalue of the covariance matrix is equal to 1. We employ a Bayes classifier to solve the corresponding classification problem.

a. Do you expect this to be a linear or non-linear classifier? Why?
b. Is this classifier equivalent to a naïve Bayes classifier? Why?
c. From the diagram, read off the eigenvectors and eigenvalues of the covariance matrix for each of the classes.
d. For each class, use the eigenvalues λ_1, λ_2 and the corresponding column eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ (normalized to unity) to reconstruct the covariance matrix Σ. If you wish, you can use the formula

$$\Sigma=\lambda_1 \boldsymbol{v}_1 \boldsymbol{v}_1^T + \lambda_2 \boldsymbol{v}_2 \boldsymbol{v}_2^T$$

e. Calculate the separating curve for the two classes.
f. Classify point $\boldsymbol{x}_0 = [3 \;\; 1]^T$.



Question 4

Probability density functions are shown for two classes *A* (red curve) and *B* (blue curve). The two probability density functions are derived from mixtures of Gaussians and are properly normalized. The two classes are equiprobable ($P(A)=P(B)=0.5$).

a. Let R_A be the union of the regions of the x axis which are assigned to class *A* by a standard Bayes classifier. Similarly, let R_B be the union of the regions which are assigned to class *B*. Show these regions on the diagram.

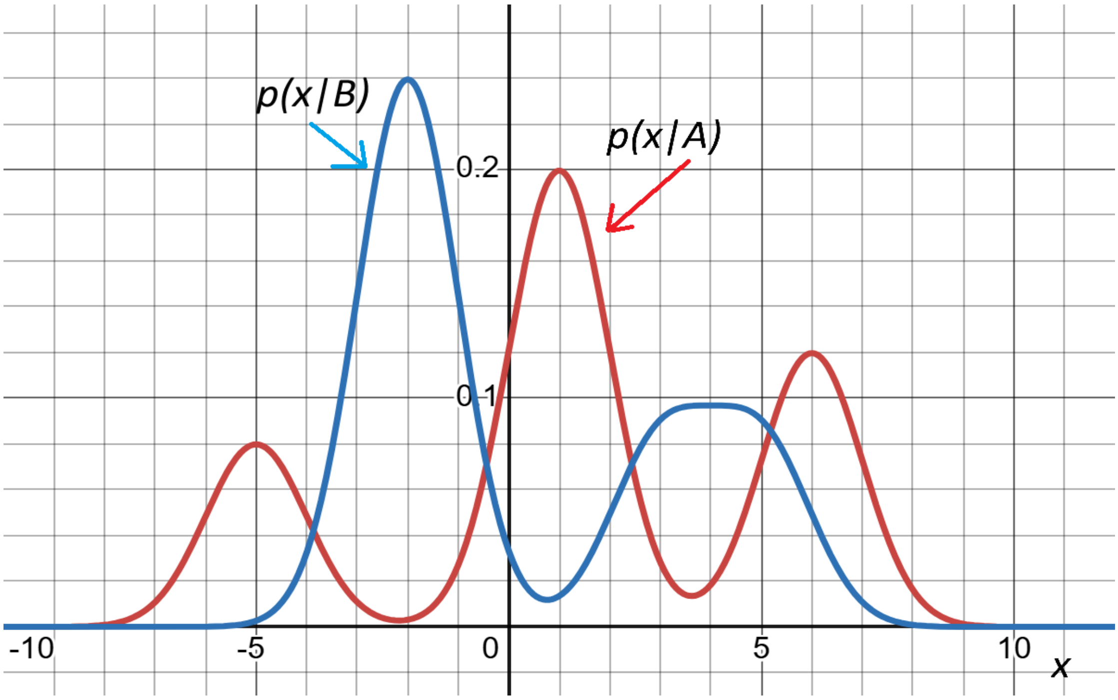
b. Shade the areas under the red curve $p(x|A)$ which correspond to regions belonging to R_B. Similarly, shade the areas under the blue curve $p(x|B)$ which correspond to regions of the x axis belonging to R_A.

c. Remember that the total probability of error is given by

$$P_e = P(A)\int_{R_B} p(x|A)dx + P(B)\int_{R_A} p(x|B)dx$$

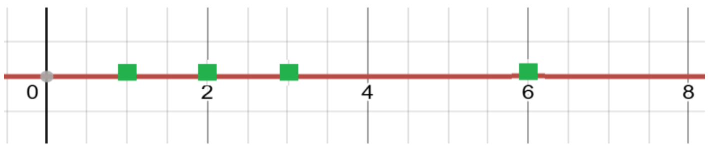
Taking into account the graphical work you did for question 4b, give a rough numerical estimate of the total probability of error.

d. Answer question 4a again, when it is considered doubly damaging to misclassify instances belonging to class *B*, relative to misclassifications of instances in class *A*.



Question 5

In a machine learning problem, our task is to estimate a parameter θ. We have the opportunity to form 4 different training sets of equal size. The estimates we get for θ from our training sets are shown in the diagram. We assume that the distribution of the results is Gaussian.



a.

a. Estimate the mean μ and standard deviation σ of the Gaussian using the Maximum Likelihood method. You are not required to go through the maximization process, just use the formulas we have already derived in class for the Gaussian distribution.

b. The true value for θ is 2. Based on the information you got from the 4 training sets evaluate the squared bias, variance and mean squared error of the estimator for θ.

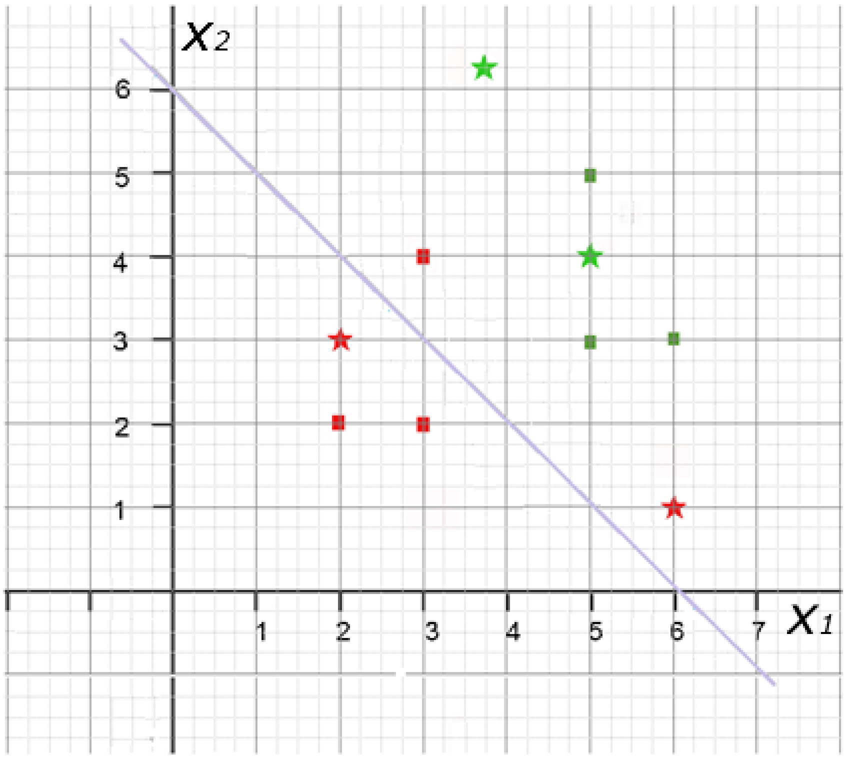
c. We suspect that we accidentally got rather large estimates for θ from our specific training sets, and therefore we now impose a prior on the mean of the Gaussian in the form of an exponential distribution with pdf equal to $p(\mu) = (1/\mu_0) \exp(-\mu/\mu_0)$ for positive μ. μ_0 is taken equal to 1. Reevaluate the mean of the Gaussian distribution using this prior and the MAP method. How much does this reduce your evaluation for the squared bias?

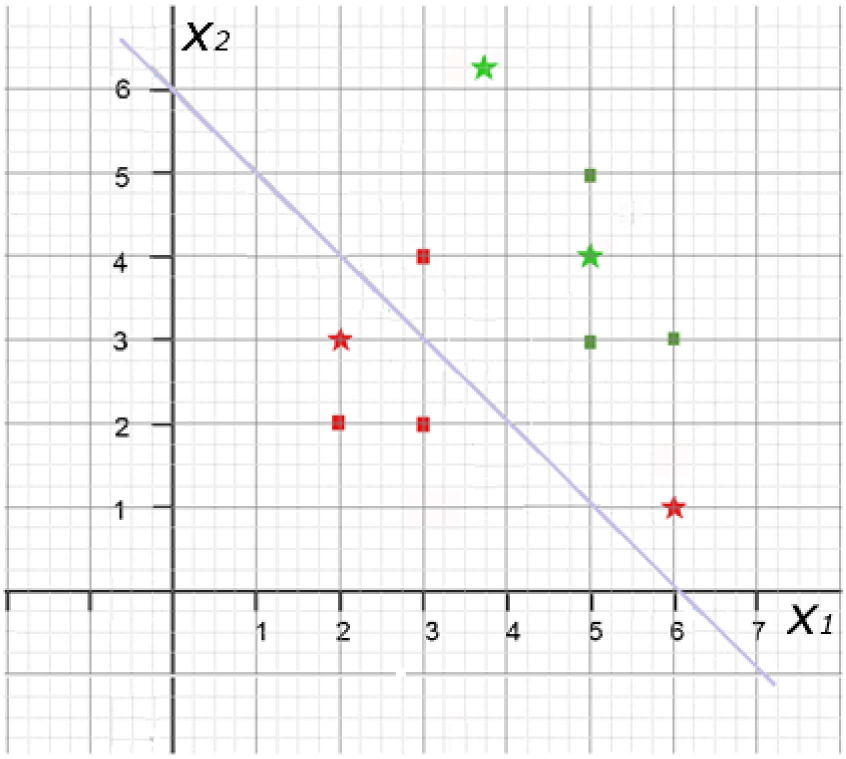
Question 6

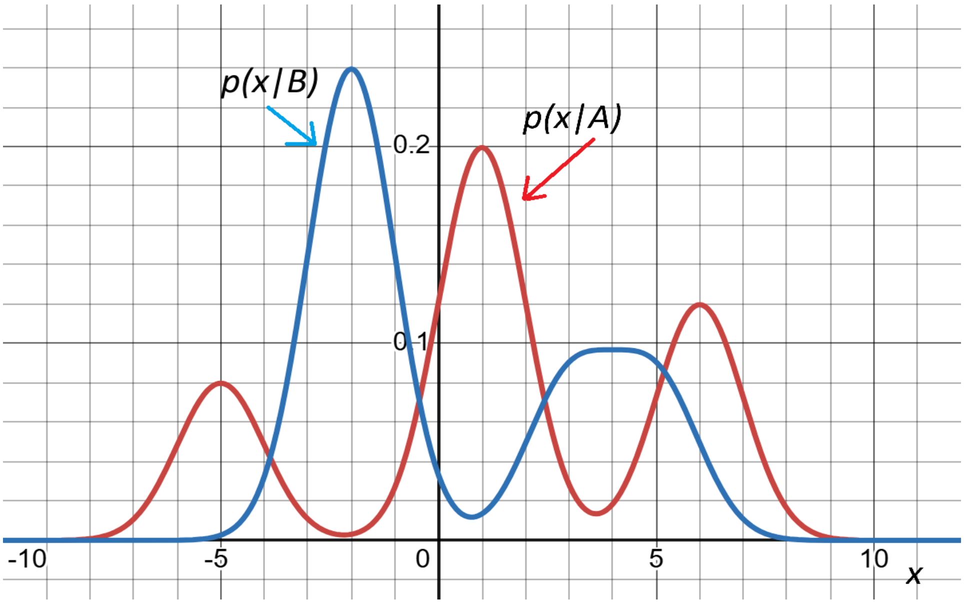
What is your motivation for studying Machine Learning in relation to your professional life and career? In what ways do you expect the social changes brought about and the ethical considerations raised by the advancement and proliferation of AI/Machine Learning to affect you?

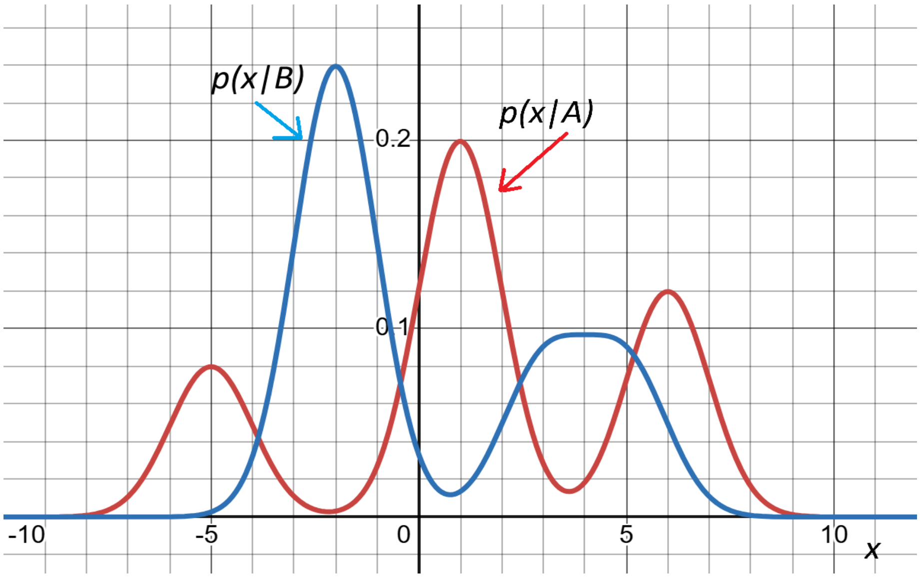
END OF EXAM

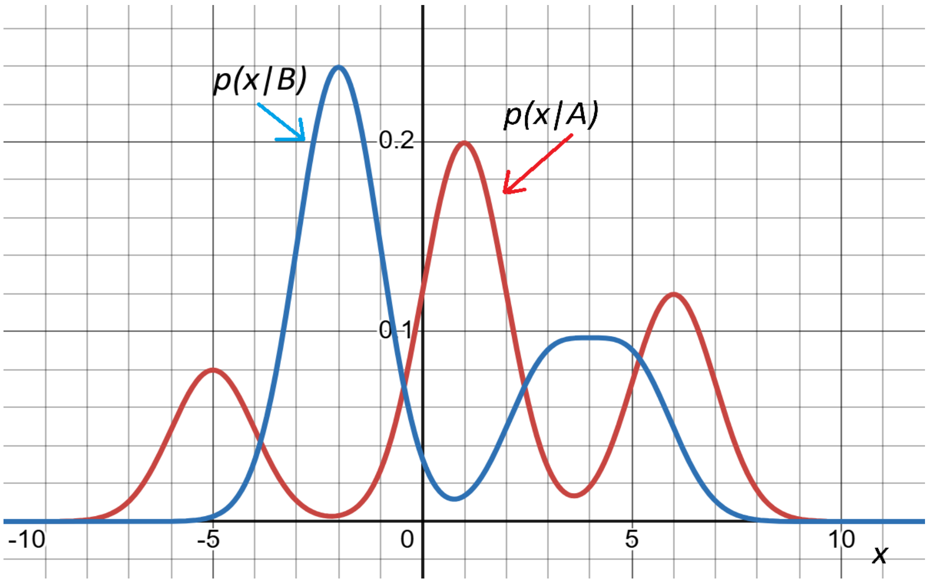
Additional copies of the diagrams






p(x|B)
p(x|A)
0.2
0.1
-10
-5
0
5
10
x


p(x|B)
p(x|A)
0.2
0.1
-10
-5
0
5
10
x


p(x|B)
p(x|A)
0.2
0.1
-10
-5
0
5
10
x