Machine Learning CS539
Midterm Exam
Spring 2022

Total score: 125 points (25 bonus points), maximum 100 points
Number of questions: 6
Number of pages: 7
Please return your response by Friday, March 4th.
Please return your response in one pdf file.
Please make sure your response to questions is written clearly.
For the application questions, you need to include your code in the pdf file.

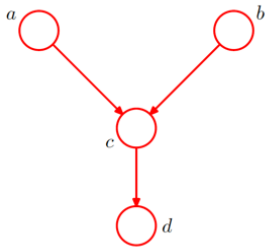
Question 1: Random variable X follows a normal distribution with mean 0 and variance σ^2 – $X \sim N(0, \sigma^2)$. **(15 pts)**

a. Let's assume $Y = |X|$. Find the pdf of Y. **(5 pts)**

b. Find the mean and variance of Y. **(5 pts)**

c. Let's assume we observe N samples from Y – in other words, rather than observing X, we observe its absolute values. Given $y_1, y_2, \ldots, y_N$, find MLE of σ^2. **(5 pts)**

Question 2: Consider the directed graph shown in the below figure in which none of the variables is observed. **(20 pts)**



a. Show that $a \perp\!\!\!\perp b \mid \emptyset$. **(5 pts)**

b. Suppose we now observe the variable d. Show that in general $a \not\perp\!\!\!\perp b \mid d$. **(5 pts)**

Let's assume a, b, and c are binary RVs, where they are taking two values 0 and 1.

$$p(a = 0) = p(b = 0) = 1/2$$

$$p(c = 0|a = 1, b = 1) = 0.1$$
$$p(c = 0|a = 0, b = 0) = 0.1$$
$$p(c = 0|a = 1, b = 0) = 0.3$$
$$p(c = 0|a = 0, b = 1) = 0.5$$

d is a continuous RV with a normal distribution. The mean of d is a function of c and its variance is 1. RV d conditional distribution is defined by

$$p(d|c) = \mathcal{N}(1 - 2c, 1)$$

c. What is the probability of c to be 1 given d is 2, a is 0, and b is 1. **(5 pts)**

d. What is the probability of c to be 1 given d is 2. **(5 pts)**

Question 3: Elastic Net (EN) is a form of regularized regression model, which is defined by

$$L(\lambda_1, \lambda_2, \mathbf{w}) = (\mathbf{t} - \mathbf{X}\mathbf{w})^{\mathrm{T}}(\mathbf{t} - \mathbf{X}\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2$$

where, λ_1 and λ_2 are positive numbers. $\mathbf{w}$ is the weight vector, $\mathbf{t}$ is the response vector, and $\mathbf{X}$ is the design matrix. **(25 pts)**

a. Write the Bayesian equivalent of EN. In other words, define the conditional distribution of the response given predictors and weight vector, and the prior model which is defined over $\mathbf{w}$. **(5 pts)**

b. Derive a recursive update rule for $\mathbf{w}$ given $\mathbf{t}$ and $\mathbf{X}$. You might require using iteratively reweighted least squares idea we discussed in HW2 (A similar idea was taken in https://web.stanford.edu/~boyd/papers/pdf/rwl1.pdf, check equations 6-7) or other possible solutions you want to pick. Provide a step-by-step explanation of your recursive algorithm. **(5 pts)**

c. Create 100 sample data points for $t_i = 1 + \underbrace{0.2}_{w_1} * x_i - \underbrace{1}_{w_2} * x_i^2 + \varepsilon_k$ where ε_k has normal distribution with a mean zero and variance of 0.1. Show how the estimated weights (w_1 and w_2) will change as a function of λ_1 and λ_2. For x, you can draw 100 random values from a normal distribution with mean 0 and variance 1. You can show your results by intensity images, where you plot w_1 (and w_2) for different values of λ_1 and λ_2. For λ_1 and λ_2, use 10 evenly spaced between 0 and 10^2 in log-scale. **(10 pts)**

d. Discuss your results in part **c** for different values of λ_1 and λ_2. **(5 pts)**

Question 4: Consider a density model given by a mixture distribution

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|k)$$

and suppose that we partition the vector $\mathbf{x}$ into two parts so that $\mathbf{x} = (\mathbf{x}_a, \mathbf{x}_b)$. **(10 pts)**

a. Show that the conditional density $p(\mathbf{x}_b|\mathbf{x}_a)$ is itself a mixture distribution. **(5 pts)**

b. Find expressions for the mixing coefficients and for the component densities. **(5 pts)**

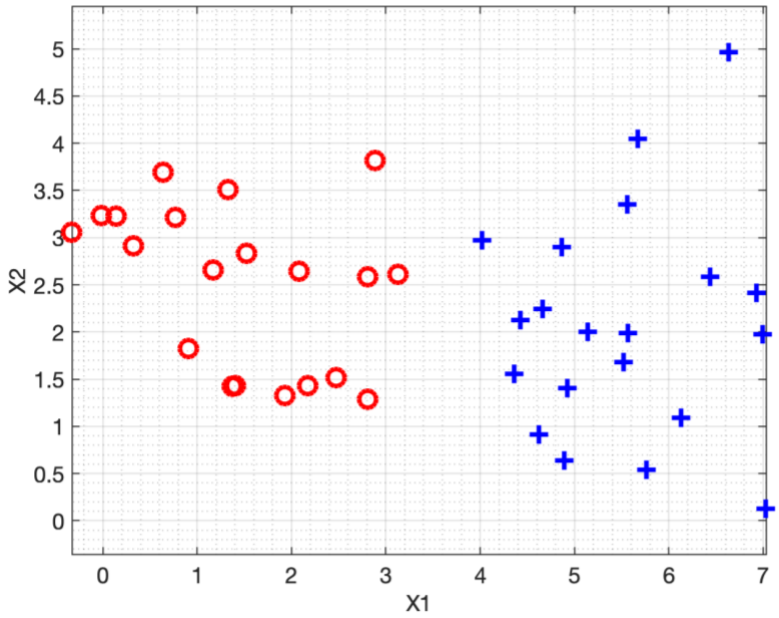
Question 5: Suppose we have survival times $x_1, x_2, x_3, \ldots, x_n \sim \text{Exponential}(\lambda)$. However, the data are being clipped (censored) if they are above c. Thus, we observe data as $(\min(x_i, c), z_i)$, where z_i is 0 when the data is $x_i \leq c$ and z_i is 1 when the data is censored. **(20 pts)**

a. Show the graphical model describing the process including z_i and survival time (x_i) and model parameter(s). **(5 pts)**

b. Find the conditional probability for $p(X = x | X \geq c)$. **(5 pts)**
Hint: here, you can check the memoryless property of exponential distribution.

c. Use EM algorithm to find MLE for λ. **(10 pts)**
Hint: here, you need to find expected values of survival times when they are censored.

Question 6: The below figure shows data samples for two classes of the data, where we have two features x_1 and x_2 per each sample of data. **(35 pts)**



a. Let's assume we fit a logistic regression model to this dataset. What would be the decision boundary, $\mathrm{w}^{\mathrm{T}}\phi(\mathrm{x}) = 0$ – note that, $\phi(\mathrm{x})$ is $[1 \;\; x_1 \;\; x_2]^{\mathrm{T}}$. Show the decision boundary in the above graph and discuss possible values for w elements. **(5 pts)**

b. A researcher suggests fitting the logistics regression with $\phi(\mathrm{x}) = [x_1 \;\; x_2]^{\mathrm{T}}$. Discuss whether the estimated decision boundary will be different from part **a** or not? Show this decision boundary (with another color) on the above graph and discuss possible values for w elements. **(5 pts)**

c. What are (expected) evidence of two models we fit to the data shown in the above figure. Discuss which one has higher evidence. Note that we have different numbers of freedom (parameter) in these two models. **(5 pts)**

d. For the dataset in the attached excel file (training.xlsx), provide model estimates for two models discussed in parts **a** and **b**. We assume a flat prior. In other words, we want to derive MLE estimate for the model parameters. In the data file, the first two columns are X1 and X2, and the last column is the class label – one and two. **(5 pts)**

e. Provide models' evidence **(5 pts)**

f. Show the model performance in the training (training.xlsx) and the test file (test.xlsx). You can create the confusion matrices, one for the training and one for the test. We will have 4 confusion matrices, 2 per each model. **(10 pts)**