EEL5840 Fundamentals of Machine Learning
Fall 2024
Midterm Exam
October 21, 2024
Time Limit: 2 hours Name: ____________________

- Write legibly
- There are a total of 8 questions for a total of 100 points
  - Some questions are worth more than other questions.
- **Closed-book, no computer, one-page formulas, calculator**
  - **Write your name in the formula sheet.**

> Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.

Helpful Formulas

- The D-dimensional **multivariate Gaussian** probability density function on random variable X, with mean μ and covariance Σ, is defined as:

$$f_X(x) = (2\pi)^{-D/2}|\Sigma|^{-1/2} \exp\left\{-(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

Closed-book, no computer, one-page formulas, calculator

Grade Table (for the teaching team use only)

Question:	1	2	3	4	5	6	7	8	Total
Points:	10	5	5	10	20	5	25	20	100
Score:									

1. (10 points) Suppose that you are training a linear regression model for a training set with N data points $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}$, and its corresponding target labels $\{t_i\}_{i=1}^N$. Consider the feature representation $\phi(x_i) = \left[x_i^2, e^{x_i}, \sin(2\pi x_i), \cos(2\pi x_i), \log(1+x_i), \frac{1}{1+x_i}\right]^T$. Answer the following questions:

   (a) (3 points) Write down the mapper function.

   (b) (3 points) Suppose you want to minimize the absolute error with the Lasso regularizer. Write down the objective function.

   (c) (4 points) What is the Bayesian interpretation of this objective function? Show and justify your work.

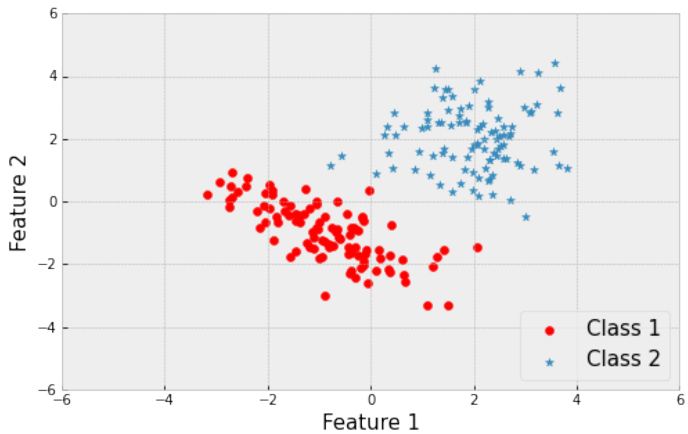
2. (5 points) What is the role of cross-validation in machine learning?

3. (5 points) Suppose that you are working with a two-dimensional features space, where x_1 and x_2 are the two features. Upon receiving a sample $\mathbf{x} = [x_1, x_2]^T$, the goal is to predict a continuous value t. Assume that you have examples of training pairs $\{(x_1, x_2)_i, t_i\}_{i=1}^{N}$. Suppose we want to train a quadratic mapper function of the form,

$$f(x_1, x_2) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 (x_1)^2 + w_5 (x_2)^2$$

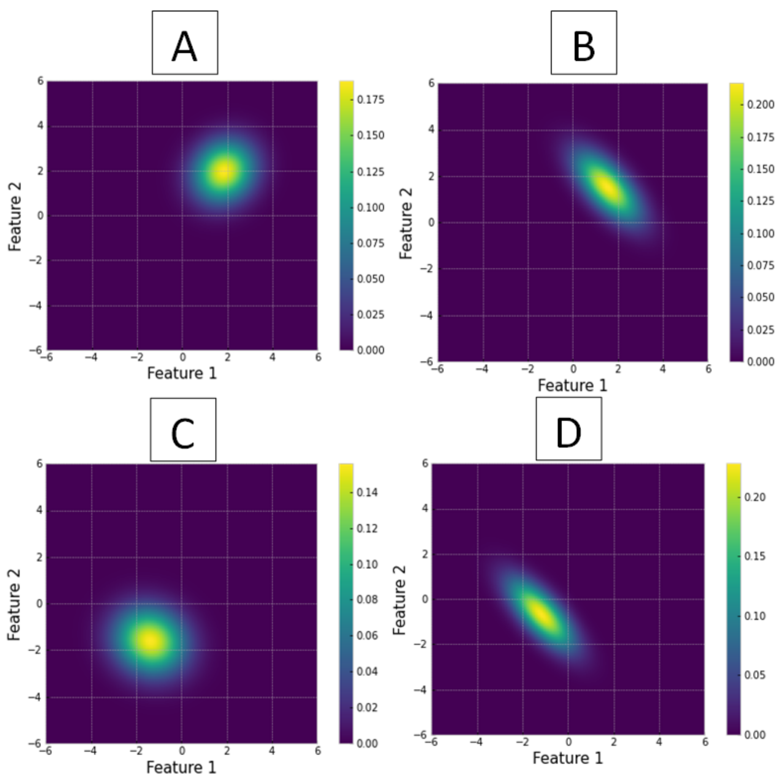
Provide a solution for $\mathbf{w} = [w_0, w_1, w_2, w_3, w_4, w_5]^T$.

4. (10 points) Consider the dataset depicted below composed of two classes (class 1 and class 2) in a 2-dimensional feature space. Each class has 100 samples. This is what the data looks like:



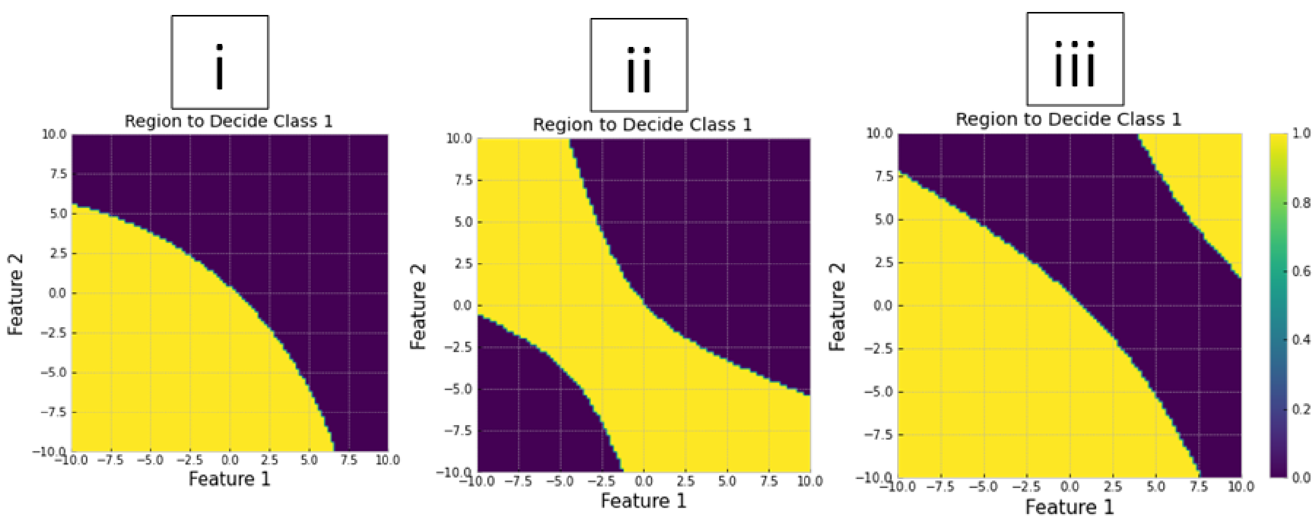
Answer the following questions:

(a) (3 points) The four plots below (A, B, C and D) represent contours plots of the data likelihood as modeled with a Gaussian distribution. From the four plots, select two plots: one that represents the data likelihood for class 1, and another that represents the data likelihood for class 2. Justify your answer.



(b) (4 points) In order to train the Naïve Bayes Classifier for this dataset, what other information do you need obtain in order to be able to make predictions? Provide your estimates below.

(c) (3 points) The three plots below (i, ii and iii) represent the decision surface for deciding class 1 (red class). Based on data likelihoods you selected in part a, which of the following plots corresponds to the decision surface for deciding class 1? Justify your answer.



5. (20 points) Suppose you have a training set with N data points $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^+$ (set of positive real numbers). Assume the samples are independent and identically distributed (i.i.d.), and each sample is drawn from a Gamma random variable with probability density function:

$$p(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where $\alpha, \beta > 0$.

Moreover, consider another Gamma density as the prior probability on the hyperparameter β,

$$p(\beta|a,b) = \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta}$$

where $a, b > 0$.

Answer the following questions:

(a) (5 points) Derive the maximum likelihood estimate (MLE) for the parameter β. Show your work.

(b) (5 points) Derive the maximum a posteriori (MAP) estimate for the parameter β. show your work.

(c) (5 points) Is the Gamma distribution a conjugate prior for the parameter β, of the Gamma data likelihood distribution? Why or why not?

(d) (5 points) Suppose you would like to update the Gamma prior distribution and the MAP point estimation in an online fashion, as you obtain more data. Write the pseudo-code for the online update of the prior parameters. In your answer, specify the new values for the parameters of the prior.

6. (5 points) Let θ_{MLE} and θ_{MAP} be the parameter estimates using MLE and MAP, respectively. In what scenarios do we achieve $\theta_{\text{MLE}} = \theta_{\text{MAP}}$?

7. (25 points) Consider a training set containing positive real numbers ($x \in \mathbb{R}^+$) for 2 classes, C_0 and C_1. The training set has 400 samples for class C_0 and 100 for C_1.

   Suppose that you have reason to believe that samples belonging from C_0 are drawn from an Exponential random variable with parameter $\lambda > 0$, and samples belonging to C_1 are drawn from a Gamma random variable with parameters $\alpha > 0$ and $\beta > 0$. In other words:

$$p(x|C_0) = \lambda e^{-\lambda x}$$
$$p(x|C_1) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

   where $\Gamma(x) = (x-1)!$. Answer the following questions:

   (a) (5 points) For a given training set $\{(x_i, t_i)\}_{i=1}^N$ with $x_i \in \mathbb{R}^+$ and $t_i \in \{0, 1\}$, how would you determine the parameters for each classes' data likelihood (λ, α and β)? Explain in words a list of steps for finding the point estimates. (No derivations.)

   (b) (5 points) Can you regularize your point estimates for λ, α and β? If yes, clearly explain how you would proceed. If not, explain why not.

(c) (10 points) For next two parts, consider $\lambda = 2$, $\alpha = 2$ and $\beta = 1$. Consider the test point $x = 3$. Which class (C_0 or C_1) will it be assigned to? Show your work.

(d) (5 points) For a given test sample x, provide an equation that will determine all cases in each x will be assigned to C_0. Show your work.

8. (20 points) Consider the dataset $\mathbf{X} = \{x_i\}_{i=1}^N$, where $x_i \geq 0, \forall i$. Suppose your goal is to perform density estimation using a Mixture Model, in particular, a Rayleigh Mixture Model. Its data likelihood can be written as:

$$f(x) = \sum_{k=1}^{K} \pi_k g_k\left(x|\sigma_k\right)$$

where

$$\sum_{k=1}^{K} \pi_k = 1$$

and

$$g_k\left(x|\sigma_k\right) = \frac{x}{\sigma_k^2} e^{-x^2/(2\sigma_k^2)}$$

with $\sigma_k > 0$ and $0 \leq \pi_k \leq 1, \forall k$. Answer the following questions:

(a) (2 points) Assuming your data is i.i.d., write down the observed data likelihood, $\mathcal{L}^0$.

(b) (2 points) Use the Expectation-Maximization (EM) algorithm to find the maximum likelihood solutions. Start by introducing the hidden latent variables Z. Describe precisely what they are.

(c) (3 points) For the hidden variables you defined above, write down the complete data likelihood, $\mathcal{L}^c$.

(d) (4 points) Write down the EM optimization function, $Q(\Theta, \Theta^t)$, where $\Theta = \{\pi_k, \sigma_k\}_{k=1}^{K}$. Your final solution should contain the sum of simple (natural-)log-terms.

(e) (4 points) Derive the update equations for the parameters σ_k.

(f) (5 points) Derive the update equations for the parameters π_k.

HONOR STATEMENT

I understand that I am bound to uphold the honor code of the University of Florida. I have neither given nor received assistance on this examination. In addition, I did not use any outside materials on this exam other than the one page of formulas that was allowed.

Sign Your Name: ______________________________

Write the Date: ______________________________

Print Your Name: ______________________________

Turn in your formula sheet with your exam!!!