CS461 Midterm II (March. 13, 2025)

CS461 Section #:	
Name:	
NetID:	

0. (3p/each) True/ False Questions

Part1:

- The posterior probability $P[A|B]$ is sensitive to the prior probability $P[A]$. (True / False)
- Maximum likelihood estimation is sensitive to the data size. (True / False)
- In an over-determined system $Ax = y$, if y is on the column space of A, a solution always exists. (True / False)
- By increasing the amount of training data, we can reduce the bias of a model. (True / False)
- Ridge regularization (adding $\lambda||w||^2$ to an objective) can be interpreted as solving an optimization problem with inequality constraints. (True / False)

Part2

- Logistic regression algorithm does not converge if data is not linearly separable. (True/ False)
- MAP classification minimizes the error rate. (True/False)
- Both Naive Bayes and logistic regression models learn posterior density $P(C_k|x)$. (True / False)
- In perceptron, when data is linearly separable, depending on the initial point, the convergence point will be different. (True / False)
- Even with an imbalance between positive and negative samples, accuracy is not affected, as it accounts for both false negatives and false positives. (True/ False)

Part3

- In general, SVM exhibits higher sensitivity (or high variance) to different training data compared to other algorithms as kernel methods like Gaussian map data into an infinite-dimensional feature space to construct a maximal margin classifier. (True/ False.)
- In SVM, removing non-support vectors does not change the decision boundary. (True /False)
- When data is not linearly separable then no way for a hard margin classifier to converge. (True /False)
- When data is linearly separable, soft and hard margin SVM will result in the same classifier. (True / False)
- A soft-margin SVM generally results in a larger margin than a hard-margin SVM. (True / False)

2.1(10p) [Bayes Rule] Before going on vacation, you ask your friend to water your ailing plant. Without water, the plant has an A percent chance of dying. Even with proper watering, it has a B percent chance of dying. And the probability that your friend will forget to water it is C percent. If the plant did not survive in the week, what is the probability that your friend watered it? (write your solution using A, B, C.)

2.2(10p) [KKT conditions] Compute the optimal solution x_1^*, x_2^*, x_3^* for the problem below. Please show all intermediate steps clearly.

$$\begin{aligned} \min \quad & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{subject to} \quad & x_1 + x_2 + x_3 \leq -3 \end{aligned}$$

3 [Linear Regression] Suppose you are given a set of data $\{(x_i, y_i) \mid i = 1, 2, \ldots, n, x_i \in \mathcal{R}, y_i \in \mathcal{R}\}$. You know that the data is artificially generated by a polynomial $f(x)$ but don't know the exact degree of it. The data contains an additive noise ϵ and it follows Gaussian.

$$y_i = f(x_i) + \epsilon_i \mid i = 1, 2, \ldots, n,$$
$$\epsilon_i \sim \mathcal{N}(0, \sigma^2 = 1.0e^{-3}) \quad \forall i$$

3.1(5p) Suppose you trained a MMSE linear regression using the basis functions of $[1, x]$ and N data ponts. The Mean Square Error (MSE) for train and test are shown below. What is a possible reason for the large MSE values?

train	MSE$\gg 1.0e-3$
test	MSE$\gg 1.0e-3$

3.2(5p) Suppose you trained a MMSE linear regression using the basis functions of $[1, x, x^2, x^3, x^4]$ and the same number of data points N in problem 3.1. The Mean Square Error (MSE) for train and test are shown below. Based on values, what issue can you identify?

train	≈ 0
test	MSE$\gg 1.0e-3$

3.3(5p) What will be two possible solutions to address the issue in problem 3.2?

3.4(5p) Suppose data is generated without intrinsic error ϵ and the polynomial is $f(x) = x + x^2 + x^3 + x^4 + x^5$. Then, how many data points will be sufficient to recover the original polynomial using MMSE linear regression? Please provide a brief explanation for your reasoning.

4. [Linear Classification] Suppose you train a Gaussian discriminant classifier based on the six data points below. When assuming equal variances for the two classes, the binary decision rule of GDA is formulated as follows.

$$P[x|C_+] \cdot P[C_+] \overset{C-}{\underset{C+}{\lesseqgtr}} P[x|C_-] \cdot P[C_-]$$

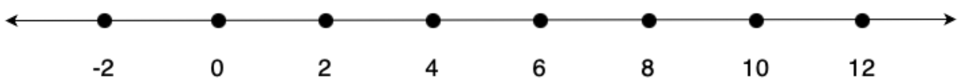
$$\leftrightarrow -\frac{1}{2}(x-\mu_+)^2 \overset{C-}{\underset{C+}{\lesseqgtr}} -\frac{1}{2}(x-\mu_-)^2, \text{ where } P[C_+] = P[C_-] = \frac{1}{2} \text{ and } \sigma_+ = \sigma_- = \sigma$$

$$\leftrightarrow x \overset{C-}{\underset{C+}{\lesseqgtr}} \frac{1}{2} \cdot (\mu_- + \mu_+)$$

4.1(5p) Given the six data points below, is the data linearly separable?

data num	x	class (t)
d_1	-10	-1
d_2	0	-1
d_3	$+10$	-1
d_4	$+11$	$+1$
d_5	$+12$	$+1$
d_6	$+13$	$+1$

4.2(5p) Compute a decision boundary and region. Please draw them below based on the MAP rule provided above.



4.3(5p) The decision boundary in problem 4.2 separates the train data? What is the margin of the classifier? hint: the margin is the smallest value of $t_i(w \cdot x_i + b)$ for data d_i.

4.4(5p) Suppose perceptron and logistic regression are trained based on the same data above. Do you think the classifiers linearly separate the training data?

	separation	no separation
Perceptron		
Logistic Regression		

4.5(5p) In class, we studied the loss of a single data point for perceptron and logistic regression, as shown in the figure below. Based on the margin in problem 4.3 and the figure, which classifier will yield the largest margin among perceptron, logistic regression, and GDA? And, which classifier has the second largest? Please provide a brief explanation for your reasoning.

5. [SVM] Suppose you train a soft-margin SVM and formulate the dual problem using the data in the table below. The lambda values in the table are the Lagrangian parameters obtained through SMO (Sequential Minimal Optimization) with $C = 1$.

$$\lambda*_{n=1}^{4} = \underset{\lambda_{n=1}^{4}}{\arg\max} -\frac{1}{2}\sum_{n=1}^{4}\sum_{m=1}^{4} \lambda_n\lambda_m \cdot t_n \cdot t_m \cdot \kappa(x_n, x_m) + \sum_{n=1}^{4}\lambda_n$$

$$\text{subject to} \quad 0 \leq \lambda_n \leq C \quad n = 1, 2, 3, 4$$

$$\sum_{n=1}^{4} \lambda_n \cdot t_n = 0$$

$$** \text{ kernel function} \quad \kappa(x_n, x_m) = x^t \cdot x$$

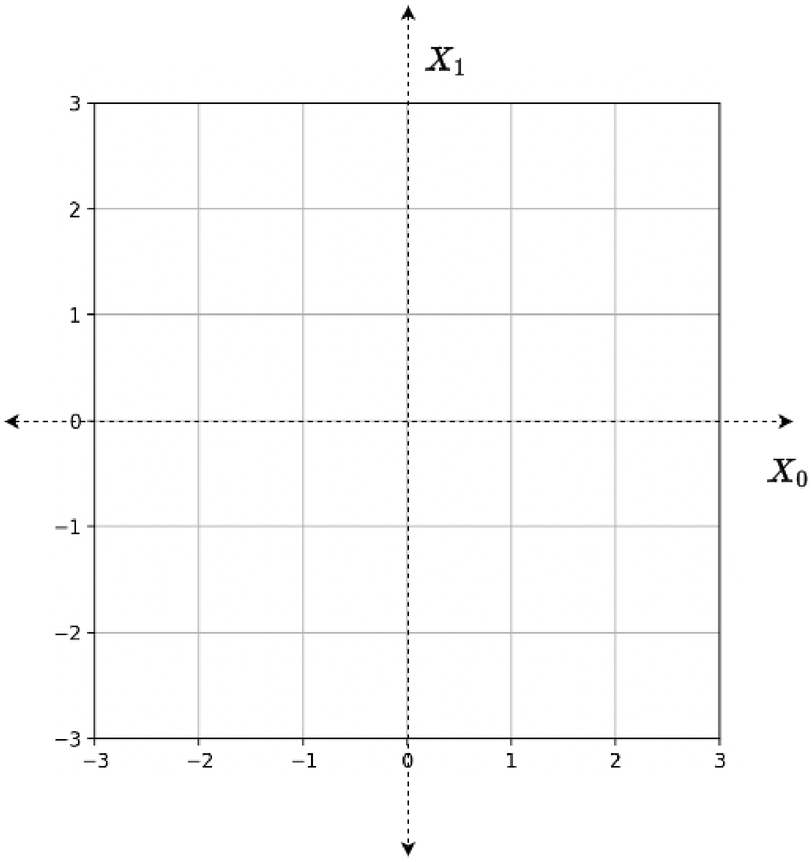
data num	$\vec{x} = (x_0, x_1)$	class	λ
d_1	$(+1, -1)$	$t_1 = +1$	$\lambda_1 = 0.64$
d_2	$(+2, -2)$	$t_2 = +1$	$\lambda_2 = 0$
d_3	$(-2, +2)$	$t_3 = -1$	$\lambda_3 = 0$
d_4	$(-1/4, +1/4)$	$t_4 = -1$	$\lambda_4 = 0.64$

5.1(5p) Based on the table, compute the decision boundary for a soft-margin SVM. Assume the bias is -0.6.

$$\delta(\vec{x}) = \sum_{n=1}^{4} \lambda_n t_n x_n^t \cdot \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} + b = 0$$

5.2(10p) Compute two margins (+ and −). Draw the decision boundary in problem 5.1, the two margins, and the four data points on the graph below.

- draw the decision boundary, margin, and four data samples.



5.3(5p) Suppose you trained the SMO with $C << 1$, where C is much smaller than 1. How would the margin differ from that of the SVM with $C = 1$?