COMP5328/COMP4328/COMP8328 Sample Final Exam 2025 Semester 2

Please help complete the USS survey (https://student-surveys.sydney.edu.au/s/)!

Question 1 [10 pts]
1). Taylor's theorem is important for optimization methods like gradient descent and Newton's method. How does Taylor's theorem help approximate future loss based on current loss?
2). What are the key differences between gradient descent and Newton's methods for update loss?

Taylor's Theorem

Let $k \geq 1$ be an integer and let the function $f : \mathbb{R} \to \mathbb{R}$ be k times differentiable at the point $a \in \mathbb{R}$. Then there exists a function $h_k : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = f(a) + f'(a)(x-a) + \ldots + \frac{f^{(k)}(a)}{k!}(x-a)^k + h_k(x)(x-a)^k$$

and $\boxed{\lim_{x \to a} h_k(x) = 0.}$

[WRITE YOUR ANSWER HERE]

Please help complete the USS survey (https://student-surveys.sydney.edu.au/s/)!

Question 2 [10 pts]

In the context of machine learning and signal processing, different norms are used to measure the sparsity of vectors. Sparsity in a vector implies that many of its elements are zero or near zero. Given the definitions of the L1, L2, and L3 norms for a vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ as follows:

- L1 norm ($\|\mathbf{x}\|_1$): Sum of the absolute values of the elements. $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- L2 norm ($\|\mathbf{x}\|_2$): Square root of the sum of the squared elements, also known as the Euclidean norm. $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
- L3 norm ($\|\mathbf{x}\|_3$): Cube root of the sum of the cubed absolute values of the elements. $\|\mathbf{x}\|_3 = \left(\sum_{i=1}^{n} |x_i|^3\right)^{1/3}$

1). Which of these norms (L1, L2, L3) is best suited for use as a surrogate for measuring the sparsity of a vector, and which is the least suitable? Provide a detailed explanation for your answer, considering using figure to explain.

[WRITE YOUR ANSWER HERE]

Question 3 [10 pts]
Consider a binary classification task where a model predicts the probability of two classes, $C = \{0, 1\}$, based on input features. You are given two data points with both clean (true) class posterior probabilities and noisy (observed) class posterior probabilities. It is assumed that the transition matrix, which represents the noise in class labels, is instance-independent (i.e., the same across all data points).

- x_1: Clean class posterior probabilities: $P(Y = 0 \mid x_1) = 0.7$. $P(Y = 1 \mid x_1) = 0.3$
  Noisy class posterior probabilities: $P(\tilde{Y} = 0 \mid x_1) = 0.5$, $P(\tilde{Y} = 1 \mid x_1) = 0.5$.
- Clean class posterior probabilities: $P(Y = 0 \mid x_2) = 0.4$, $P(Y = 1 \mid x_2) = 0.6$.
  Noisy class posterior probabilities: $P(\tilde{Y} = 0 \mid x_2) = 0.6$, $P(\tilde{Y} = 1 \mid x_2) = 0.4$.

Calculate the transition matrix with the provided information. Show your calucation step in detals.
[WRITE YOUR ANSWER HERE]

Question 4 [10 pts]

In the realm of domain adaptation, understanding and addressing covariate shift is crucial. Covariate shift occurs when the probability distributions of input data in the training (source) and testing (target) domains differ, specifically $p_s(X) \neq p_t(X)$, where $p_s(X)$ and $p_t(X)$ denote the probability densities of the source and target domains, respectively. Addressing this shift is essential for adapting models trained on the source domain to perform effectively in the target domain.

1). Explain why covariate shift can be problematic when training models.

2). Describe how importance reweighting can be used to address covariate shift, including a detailed step-by-step calculation of the weights.

3). What assumptions underlie the use of importance reweighting for covariate shift?

[WRITE YOUR ANSWER HERE]

Question 5 [10 pts] Consider a loss function $\phi(z) = \log(1 + \exp(-z))$.
1). What is the key property of a classification-calibrated loss function? How to check if a loss function is classification-calibrated?
3). Is $\phi(z) = \log(1 + \exp(-z))$ a classification-calibrated loss function? Provide a detailed explanation and show the calculation steps in detail. Note that the derivative of $\exp(x)$ with respect to a variable x is $\exp(x)$ itself, and $\exp(x) > 0, \forall x$.
[WRITE YOUR ANSWER HERE]