ECE 461/661 Spring 2025 Practice Questions for Mini-Exam 3

Introduction to Machine Learning for Engineers
Prof. Gauri Joshi and Prof. Carlee Joe-Wong

The number of practice questions included in this document may not be representative of the length of the exam. These are solely intended for you to gain more experience in answering questions similar to what you will see on the exam.

Problem 1: *[2 points]* Which of the following statements best describes the role of Transformers in machine learning?

- ◯ Transformers use convolutional layers to process sequential data efficiently.
- ◯ Transformers rely on recurrent connections to handle long-range dependencies in text.
- ◯ Transformers use self-attention mechanisms to weigh the importance of different words in a sequence.
- ◯ Transformers require labeled data for training and cannot be used in unsupervised learning.

Solution: C, Transformers use self-attention mechanisms to weigh the importance of different words in a sequence, allowing them to capture long-range dependencies effectively without relying on recurrence.

Problem 2: *[2 points]* The transformer architecture partly consists of a stack of transformer blocks. Each transformer block contains multiple components within it. Which of the following is **NOT** a component of a typical transformer block?

- ◯ Fully connected neural network.
- ◯ Attention mechanism
- ◯ Residual connections
- ◯ Positional encoding

Solution: D. Positional encoding is found only at input encoding and is not part of each transformer block.

Problem 3: *[2 points]* Which of the following statements best describes the effect of increasing the local update period τ in Local-update SGD?

- ◯ Increasing τ always improves convergence by reducing the number of communication rounds.
- ◯ A larger τ decreases the error floor by allowing more local computation.
- ◯ Increasing τ reduces communication overhead, but can hurt convergence due to divergence among local models.
- ◯ With larger τ, the local models stay better synchronized, leading to improved convergence.

Solution: C.

Problem 4: *[1 points]* K-means clustering always converges to a local optimum.

◯ True

◯ False

Solution: True.

Problem 5: *[1 points]* K-means clustering is sensitive to the initial placement of centroids.

◯ True

◯ False

Solution: True.

Problem 6: *[1 points]* K-means++ initializes centroids iteratively with probability inversely-proportional to distance from the previous center.

◯ True

◯ False

Solution: False.

Problem 7: *[1 points]* Principal Component Analysis (PCA) minimizes the reconstruction error of the data under the constraint that the projection directions are orthonormal.

◯ True

◯ False

Solution: True.

Problem 8: *[4 points]* Consider the d-dimensional data matrix with n data point represented by $\mathbf{X} \in \mathbb{R}^{n \times d}$. What is the computational complexity of PCA of $\mathbf{X}$ obtained by computing the covariance matrix $\mathbf{\Sigma} = \mathbf{X}^\top \mathbf{X}$ and computing eigenvectors of $\mathbf{\Sigma}$.

◯ $\mathcal{O}(nd^2 + d^3)$

◯ $\mathcal{O}(n^2 d + d^3)$

◯ $\mathcal{O}(nd^2 + d^2)$

◯ $\mathcal{O}(n^2 d + d^2)$

Solution: A – $\mathcal{O}(nd^2 + d^3)$