

18-661: Introduction to ML for Engineers

HW 4 Code Review and Practice Questions for Mini-Exam 3

Spring 2025

ECE – Carnegie Mellon University

HW 4, Q4: Clustering Human Activity using Inertial Sensors Data

Question: 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.

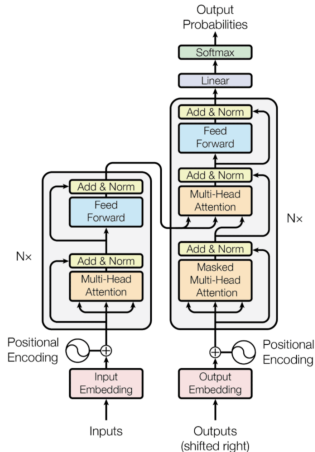
C. Transformers use self-attention mechanisms to weigh the importance of different words in a sequence, allowing them to capture long-range dependencies effectively without relying on recurrence.

Question: 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

Transformers

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



Question: 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.

C. Increasing τ reduces communication overhead, but can hurt convergence due to divergence among local models.

Question: K-means clustering always converges to a local optimum.

- True
- False

True. It always converges to a local minimum with the distortion measure monotonically decreasing to 0 at every iteration.

Question: K-means clustering is sensitive to the initial placement of centroids.

- True
- False

True. Different initializations yield in different clusterings.

Question: K-means++ initializes centroids iteratively with probability inversely-proportional to distance from the previous center.

- True
- False

False. It is directly-proportional.

Question: Principal Component Analysis (PCA) minimizes the reconstruction error of the data under the constraint that the projection directions are orthonormal.

- True
- False

True.

Dimensionality Reduction

Question: 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 $\mathbf{\Sigma}$.

- $\mathcal{O}(nd^2 + d^3)$
- $\mathcal{O}(n^2d + d^3)$
- $\mathcal{O}(nd^2 + d^2)$
- $\mathcal{O}(n^2d + d^2)$

A. $\mathcal{O}(nd^2 + d^3)$. Computing the covariance matrix takes $\mathcal{O}(d \times n \times d)$ number of operations and computing the eigenvectors of the covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$ matrix takes $\mathcal{O}(d^3)$ number of operations.