EE/CS 228 Practice Exam

Name (print):

Student ID:

Remarks:

- Any consultation with others will lead to a failing grade in the course and will be reported for disciplinary action. **If you are caught copying someone else's solution, you will FAIL.**

- This practice exam includes 7 questions. The actual midterm will 30 questions.

- Some questions might have **multiple** correct answers. You will get 1 point if you select all correct answers, 0.5 points if you select a subset of the correct answers, and 0 if your selection contains any incorrect answer.

- Please write down your answer in front of each question.

- The last 4 pages are intentionally left blank, which can be removed and used as scratch papers.

- GOOD LUCK!

Problems

Problem 1. Which statements are true about the role of non-linear activation functions in neural networks?

    (a) Non-linear activation functions are optional in deep networks, as linear transformations are sufficient for complex representations.

    (b) ✓Non-linear activations enable the network to approximate complex, non-linear functions.

    (c) ✓Without non-linear activations, a MLP with multiple layers behaves as if it only has one layer.

    (d) ✓The choice of activation function can impact the training dynamics, such as the risk of vanishing or exploding gradients.

Problem 2. Which comments are true regarding the use of Convolutional Neural Networks (CNNs) for image classification tasks?

    (a) ✓CNNs use filters to detect spatial patterns like edges and textures in images.

    (b) Pooling layers in CNNs contain learnable parameters.

    (c) Filters in CNNs are pre-defined and cannot be learned.

    (d) ✓CNNs with more layers are usually better at capturing more complex image features.

Problem 3. Which of the following statements about residual connections in deep networks are correct?

    (a) Residual connections are primarily used in recurrent networks to capture long-term dependencies.

    (b) ✓Residual connections help mitigate the vanishing gradient problem by allowing gradients to flow through identity mappings.

    (c) ✓Residual connections allow the network to learn small incremental changes from one layer to the next.

    (d) Residual connections can only be used in CNNs but not transformers.

Problem 4. Consider a network $f(x) = v^T \text{avg-pool}(x)$. Here, x is a 7×1 vector, the average-pooling operation has a receptive field of 3 and stride of 1, and v is a 5×1 vector. What is the output of the network at $x = [5, 4, 3, 2, 3, 4, 5]$ and $v = [1, 1, 1, 1, 1]$?

    (a) 20

    (b) ✓$\frac{50}{3}$

    (c) $\frac{55}{3}$

    (d) 18

Problem 5. Suppose you have a dataset $X = [x_1 \ \dots \ x_n]^T$ and $y = [y_1 \ \dots \ y_n]^T$. Consider the least squares problem $L(\theta) = \|y - X\theta\|^2$. Assume $X^T X$ is invertible. What is the minimizer of $L(\theta)$?

    (a) ✓$(X^T X)^{-1} X^T y$

    (b) $X(XX^T)^{-1} y$

    (c) $(XX^T)^{-1} X^T y$

    (d) $X(X^T X)^{-1} y$

Problem 6. Suppose you have a fully-connected network with one hidden layer. The input dimension is 300, the hidden layer has dimension 100, and the output layer is 30 dimensional. What is the total number of parameters in this network? (Please assume that there are also bias variables in each layer.)

    (a) 33000

    (b) ✓33130

    (c) 33300

    (d) 33600

Problem 7. Suppose the input tensor to a convolutional neural network is $64 \times 64 \times 3$ dimensional. The network consists of a single convolutional layer with 20 filters of size 5×5 and uses stride 2. What is the total number of parameters in this network? (Assume no bias variables.)

    (a) 1500

    (b) ✓1500

    (c) 7500

    (d) 10000

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.