Quiz 5: Regularization and CNNs

Instructions

Select the best answer for each question. Questions 1-8 cover core concepts from Session 4, while questions 9-10 are more challenging.

Questions

1. What is the primary characteristic of overfitting in neural networks?

A) The model performs poorly on both training and validation data. B) The model performs well on training data but poorly on unseen validation data. C) The model has too few parameters to capture the underlying patterns. D) The model's training loss increases over time.

2. Which of the following is NOT a common regularization technique mentioned?

A) L2 Regularization (Ridge) B) Dropout C) Feature Expansion D) Early Stopping

3. What is the main purpose of early stopping?

A) To speed up the training process significantly. B) To prevent the model from learning noise by stopping when validation performance degrades. C) To ensure the model reaches perfect training accuracy. D) To automatically adjust the learning rate during training.

4. Convolutional Neural Networks (CNNs) are particularly well-suited for what type of data?

A) Sequential data like text B) Tabular data with clear columns C) Spatial data like images D) Time-series data from sensors

5. What is the key concept behind convolutional layers in CNNs that reduces the number of parameters compared to fully connected layers?

A) Max pooling B) Residual connections C) Parameter sharing (using the same filter across the input) D) ReLU activation function

6. What is the typical function of a pooling layer in a CNN?

A) To increase the dimensionality of the feature maps. B) To introduce non-linearity into the network. C) To reduce the spatial size of feature maps and provide some translation invariance. D) To apply different filters to detect various patterns.

7. Which architecture introduced residual connections to enable much deeper networks by mitigating the vanishing gradient problem?

A) LeNet-5 B) AlexNet C) ResNet D) VGGNet

8. What is transfer learning in the context of CNNs?

A) Training a CNN from scratch on a new dataset. B) Using a pre-trained CNN model and adapting it for a new, related task. C) Converting a CNN model to run on different hardware. D) Sharing parameters between different layers within the same CNN.

9. A CNN model achieves 99% accuracy on the training set but only 75% on the validation set. Which strategy is LEAST likely to directly address this specific issue?

A) Adding more convolutional layers to increase model capacity. B) Applying dropout regularization to the fully connected layers. C) Using data augmentation techniques on the training images. D) Implementing early stopping based on validation accuracy.

10. How does parameter sharing in convolutional layers contribute to translation invariance?

A) By ensuring each output neuron is connected to every input neuron. B) By applying the same filter (pattern detector) across all spatial locations of the input. C) By normalizing the activations within each layer. D) By randomly dropping neurons during training.

Answers

1. B
2. C
3. B
4. C
5. C
6. C
7. C
8. B
9. A
10. B

Explanations

1. Overfitting is defined by the model learning the training data too well, including noise, leading to poor generalization on new, unseen data (validation set).
2. Feature expansion generally increases model complexity and is not a regularization technique; L1/L2, Dropout, Early Stopping, Batch Norm, and data augmentation are common regularization methods or strategies to combat overfitting.

3. Early stopping monitors validation performance and stops training when it begins to worsen, preventing the model from overfitting to the training data.
4. CNNs leverage spatial hierarchies and are designed primarily for grid-like data structures, making them exceptionally effective for images and other spatial data.
5. Parameter sharing (using the same filter weights across different spatial locations) and local connectivity are the core ideas in convolutional layers that dramatically reduce parameters compared to dense layers.
6. Pooling layers (like max pooling) downsample the feature maps, reducing computational cost, controlling overfitting, and providing a degree of invariance to the exact position of features.
7. ResNet (Residual Networks) introduced residual connections ($F(x) + x$), allowing gradients to flow more easily through very deep networks, overcoming the vanishing gradient problem.
8. Transfer learning involves taking a model trained on a large dataset (like ImageNet) and fine-tuning its later layers (or adding new ones) for a different, often smaller, dataset, leveraging the learned features.
9. High training accuracy and low validation accuracy indicate overfitting. Adding more layers increases model capacity, which is likely to worsen overfitting, not fix it. Regularization (B, C) and early stopping (D) are standard techniques to combat overfitting.
10. Because the same filter (weights) is applied across the entire input, it can detect the same pattern regardless of where it appears spatially in the image, contributing to translation invariance.