Quiz 5

Time limit. 30 minutes.

Collaboration and materials. You must complete this quiz individually. You may not use any materials (physical or electronic) besides both sides of one sheet of 8.5x11-inch paper with 1-inch margins and the equivalent of 10-point font.

Questions. This quiz has ten multiple-choice questions. Some questions require you to select exactly one of the answer choices, while others require you to select all of the answer choices that apply. Questions of the latter kind always end with "Select all that apply."

Scoring. Each question is weighted equally. For questions requiring you to select one of the answer choices, no partial credit will be awarded. For questions requiring you to select all of the answer choices that apply, partial credit will be awarded for each correct answer selected while no points will be awarded if no correct answers are chosen or if any incorrect answers are selected.

Submission. You will receive a bubble sheet for your answers. Please print your full name as it appears on Gradescope (please no cursive), your student ID, and today's date (November 30). You may leave the "Section" box blank. **Your version is B. Please check that this matches the pre-bubbled version number at the top of the bubble sheet.** For each question, please fill in the appropriate bubbles completely using either pencil or blue/black pen. If you have filled in a bubble with pen but have changed your mind, you can cross out that bubble with an X. Note that the answer choices are presented in the order A, B, C, D, E.

1 1 point

We train a neural network on 10,000 observations using stochastic gradient descent, with mini-batch size 200. If computing each stochastic gradient step takes 1 second, it takes N seconds to run 10 epochs. The number N has A digits, and the sum of these digits is B. What is A + B?

- 8
- 9
- 10
- 11
- 12

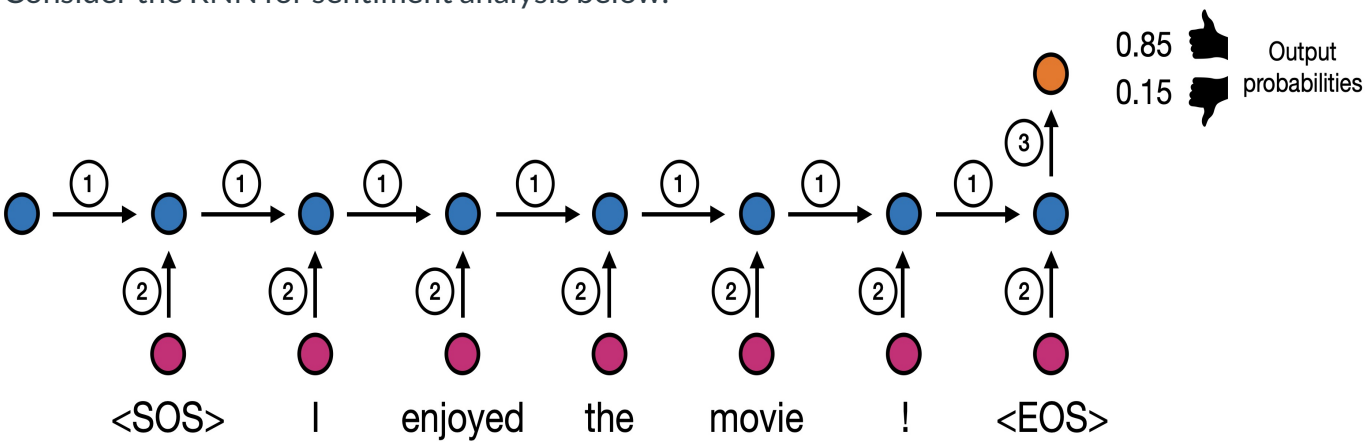
2 1 point

Consider a basic RNN to predict whether a news article is about sports, politics, entertainment, or business. It is based on word vectors and hidden state vectors of length four. This RNN has N total parameters. What is the sum of the digits of N?

- 6
- 9
- 12
- 15
- 18

3 1 point

Consider the RNN for sentiment analysis below:



Which sets of weights are used in the process of feature extraction?

- 1 only
- 1 and 2 only
- 2 and 3 only
- 1 and 2 and 3
- 3 only

4 1 point

A convolutional neural network inputs 32x32x3 images. The first convolutional layer uses 10 filters, each of size 3x3x3. The resulting activation map has dimensions WxHxD. What is the sum of the digits of the number W + H + D?

- 5
- 7
- 9
- 11
- 13

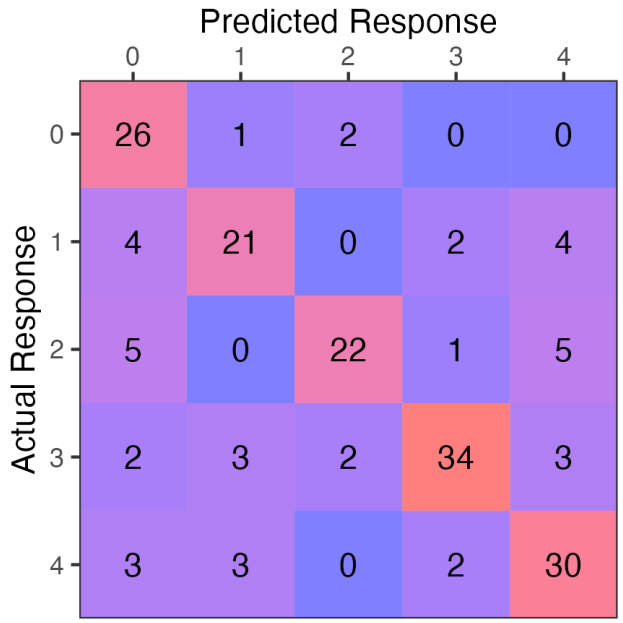
5 1 point

Consider classifying an image into one of three categories: dog, cat, or horse. Which of the following probability assignments leads to the **highest** cross-entropy loss?

- True class: Dog. Estimated probabilities: $p_{\text{dog}} = 0.1,\ p_{\text{cat}} = 0.1,\ p_{\text{horse}} = 0.8$
- True class: Cat. Estimated probabilities: $p_{\text{dog}} = 0.6,\ p_{\text{cat}} = 0.3,\ p_{\text{horse}} = 0.1$
- True class: Horse. Estimated probabilities: $p_{\text{dog}} = 0,\ p_{\text{cat}} = 0.3,\ p_{\text{horse}} = 0.7$
- True class: Dog. Estimated probabilities: $p_{\text{dog}} = 0.6,\ p_{\text{cat}} = 0.3,\ p_{\text{horse}} = 0.1$

6 1 point

A CNN was trained for the handwritten digit classification task, restricted to the digits {0, 1, 2, 3, 4}. Below is the confusion matrix of a this CNN on a test dataset. Which digit was misclassified the largest number of times? [Misclassifying the digit D means predicting a class other than D when the true digit is D.]



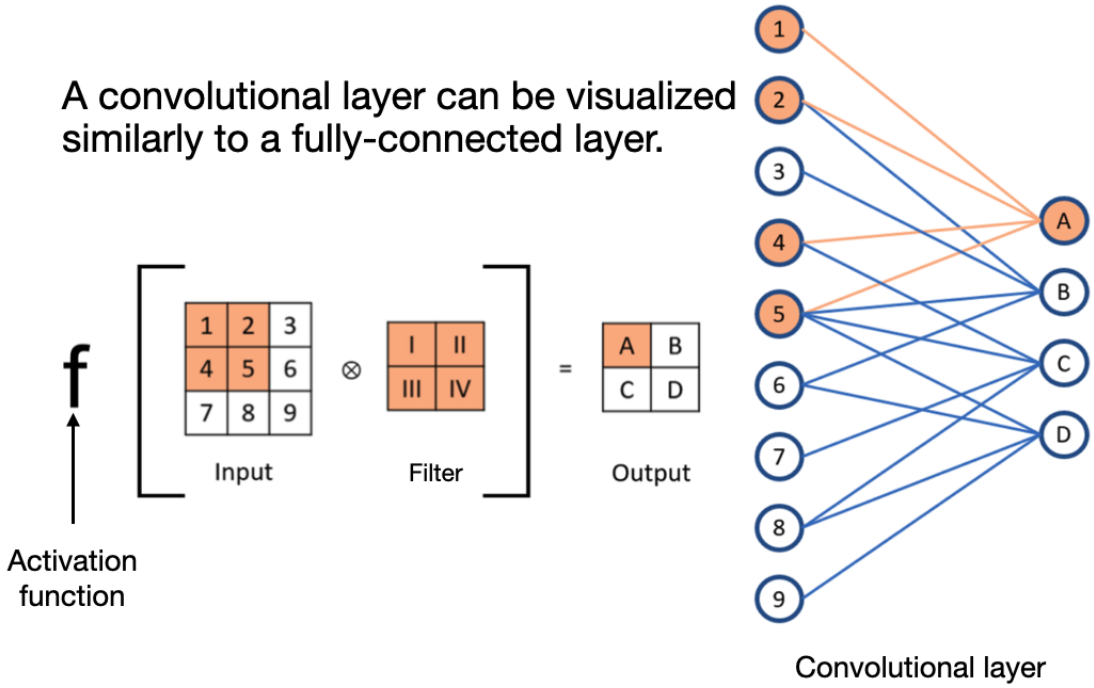
- 0
- 1
- 2
- 3
- 4

7 1 point

Which of the models below are capable of learning derived features (also known as feature learning)? Select all that apply.

- [ ] Recurrent neural network
- [ ] Convolutional neural network
- [ ] Fully connected neural network with one hidden layer
- [ ] Multiclass logistic regression
- [ ] Linear regression

8 0.5 points



Which weight connects node 5 and node B in the network at right?

- ( ) I
- ( ) II
- ( ) III
- ( ) IV
- ( ) Not enough information given.

9 1 point

Consider a multi-class logistic regression model for classifying 50x50x3 color images of animals into 10 categories. This model has N parameters. The number N has A digits, and the sum of these digits is B. What is A + B?

- 13
- 15
- 17
- 18
- 20

10 1 point

Which of the following helps neural networks avoid overfitting? Select all that apply.

- [ ] Training sets with large numbers of observations
- [ ] Using the cross-entropy loss function
- [ ] Dropout
- [ ] Using the validation set approach instead of cross-validation
- [ ] Early stopping