

Quiz 5

You have 30 minutes to complete this 10-question quiz. The questions are weighted equally. You can consult any course materials, the internet, or R. However, you must complete the quiz individually.

1 Numeric 0.5 points

Consider a multi-class logistic regression model for classifying $32 \times 32 \times 3$ color images of animals into 10 categories. The input is a flattened $32 \times 32 \times 3$ image (the last dimension corresponds to R,G,B channels) and the output is a set of 10 probabilities, one for each of the 10 categories. How many parameters does this model have?

30,730

2 Multiple Choice 0.5 points

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: $p_{\text{dog}} = 0.6$, $p_{\text{cat}} = 0.3$, $p_{\text{horse}} = 0.1$
- ☐ 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: $p_{\text{dog}} = 0.1$, $p_{\text{cat}} = 0.1$, $p_{\text{horse}} = 0.8$

Below is the confusion matrix for a neural network model evaluated on the MNIST test data. Which digit did this neural network misclassify most frequently? Note that all ten digits are equally represented in the MNIST test data.

		Predicted Response									
		0	1	2	3	4	5	6	7	8	9
Actual Response	0	965	0	1	1	0	3	7	1	2	0
	1	0	1118	3	2	1	0	3	1	7	0
	2	9	2	968	8	10	1	7	7	17	3
	3	3	1	15	950	1	11	2	12	11	4
	4	1	1	3	0	936	0	10	2	3	26
	5	7	1	2	26	6	815	12	3	14	6
	6	10	3	4	1	11	6	920	0	3	0
	7	2	9	21	6	6	0	1	959	2	22
	8	5	4	5	19	6	8	13	11	895	8
	9	11	7	1	11	34	4	1	10	3	927

5

4

Multiple Answer 0.5 points

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

- ☐ Linear regression
- ☐ Multiclass logistic regression
- ☒ Fully connected neural network with one hidden layer
- ☒ Convolutional neural network

5

Numeric 0.5 points

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, how many seconds does it take to run 10 epochs?

6

Multiple Answer 0.5 points

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

- ☐ Learning via stochastic gradient descent
- ☒ Training sets with large numbers of observations
- ☒ Dropout
- ☒ Early stopping
- ☐ Tuning via a validation set rather than cross-validation

7

Numeric 0.5 points

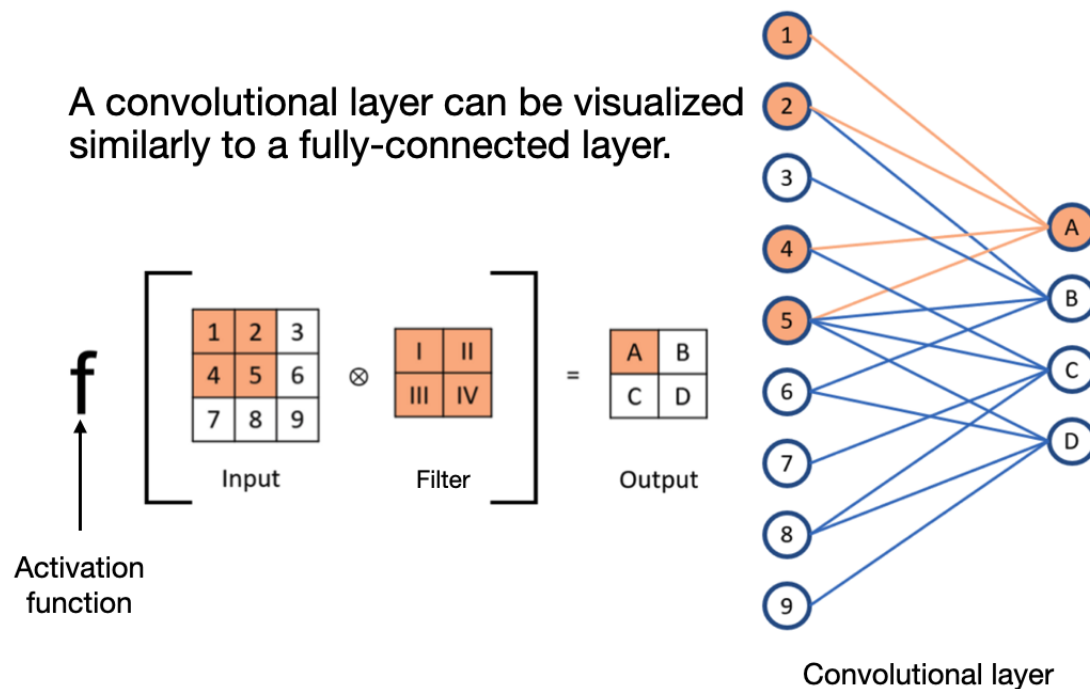
A convolutional neural network inputs $32 \times 32 \times 3$ images. The first convolutional layer uses 10 filters, each of size 3×3 . How many channels does the resulting activation map contain?

10

8

Multiple Choice 0.5 points

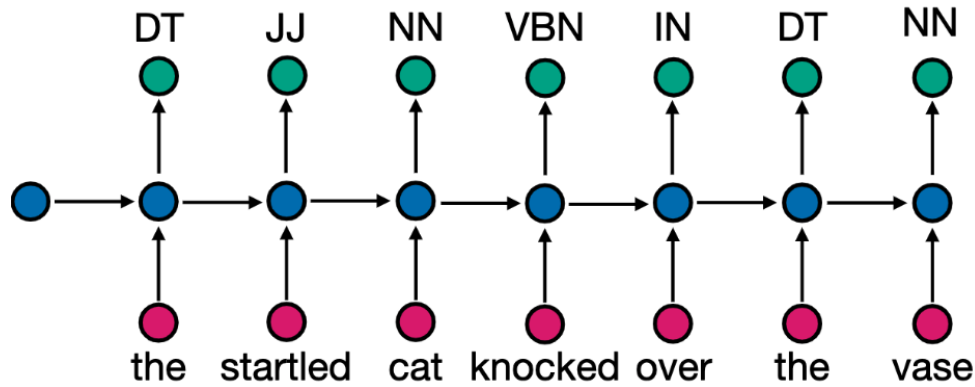
A convolutional layer can be visualized similarly to a fully-connected layer.



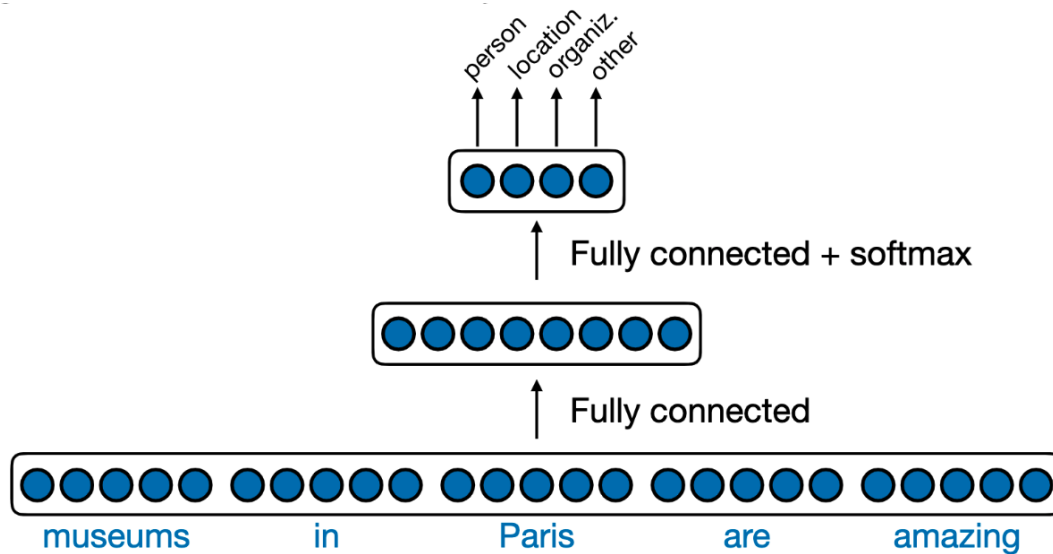
Which weight connects node 5 and node B in the network at right?

- ☐ I
- ☐ II
- ☒ III
- ☐ IV

Consider the RNN below, designed for part-of-speech tagging. Each arrow in the diagram represents a set of model parameters. How many **distinct** sets of model parameters are there to be learned?



Consider the fully-connected neural network below, used for named entity recognition. It is based on word vectors of length 5, pre-trained for a vocabulary of 10,000 words and not fine-tuned.



Which of the following statements are correct about the number of parameters in this model? Select all that apply.

- ☒ Increasing the word vector length will increase the number of model parameters.
- ☐ Removing the softmax transformation from the last layer will decrease the number of model parameters.
- ☐ Increasing the vocabulary size will increase the number of model parameters.
- ☒ Increasing the number of output categories will increase the number of model parameters.