

 Try again once you are ready

Item 1 of 10

Grade received 77% To pass 80% or higher

Try again

Practical aspects of Deep Learning

Quiz • 20 min

Practical aspects of Deep Learning

Latest Submission Grade 77%

Due Apr 4, 2:59 PM CST Attempts 3 every 8 hours

1. If you have 10,000,000 examples, how would you split the train/dev/test set? 1 / 1 point

Try again

☐ 33% train . 33% dev . 33% test

☒ Receive grade
60% train . 20% dev . 20% test

To Pass 80% or higher
☒ 98% train . 1% dev . 1% test

Your grade 77% Correct

View Feedback

2. In a personal experiment, an M.L. student decides to not use a test set, only train-dev sets. In this case which of the following is We keep your highest score 1 / 1 point

☒ He might be overfitting to the dev set.

☐ Like ☐ Dislike
He won't be able to measure the variance of the model.

☐ Not having a test set is unacceptable under any circumstance.

☐ He won't be able to measure the bias of the model.

☒ Correct
Yes. Although not recommended, if a more accurate measure of the performance is not necessary it is ok to not use a test set. However, this might cause an overfit to the dev set.

3. If your Neural Network model seems to have high variance, what of the following would be promising things to try? 0.2 / 1 point

☐ Add regularization

☒ Increase the number of units in each hidden layer

☒ This should not be selected

☐ Get more training data

☒ Make the Neural Network deeper

☒ This should not be selected

☐ Get more test data

4. Working on a model to classify bananas and oranges your classifier gets a training set error of 0.1% and a dev set error of 11%. Which of the following two are true? 0.5 / 1 point

☐ The model is overfitting the train set.

☐ The model has a very high bias.

☒ *D: The model has a high variance.

☒ Correct
No. This model has a low bias and high variance.

☒ The model is overfitting the dev set.

☒ This should not be selected
No. This would imply a very low error on the dev set.

5. What is weight decay? 1 / 1 point

☐ A technique to avoid vanishing gradient by imposing a ceiling on the values of the weights.

☒ A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration.

☐ The process of gradually decreasing the learning rate during training.

☐ Gradual corruption of the weights in the neural network if it is trained on noisy data.

☒ Correct

6. What happens when you increase the regularization hyperparameter lambda? 1 / 1 point

☒ Weights are pushed toward becoming smaller (closer to 0)

☐ Gradient descent taking bigger steps with each iteration (proportional to lambda)

☐ Weights are pushed toward becoming bigger (further from 0)

☐ Doubling lambda should roughly result in doubling the weights

☒ Correct

7. Which of the following are true about dropout? 1 / 1 point

☐ In practice, it eliminates units of each layer with a probability of keep_prob.

☐ It helps to reduce the bias of a model.

☒ In practice, it eliminates units of each layer with a probability of 1- keep_prob.

☒ Correct
Correct. The probability that dropout doesn't eliminate a neuron is keep_prob.

☒ It helps to reduce overfitting.

☒ Correct
Correct. The dropout is a regularization technique and thus helps to reduce the overfit.

8. During training a deep neural network that uses the tanh activation function, the value of the gradients is practically zero. Which of the following is most likely to help the vanishing gradient problem? 0 / 1 point

☐ Use Xavier initialization.

☐ Increase the number of layers of the network.

☒ Use a larger regularization parameter.

☐ Increase the number of cycles during the training.

☒ Incorrect
Incorrect. Regularization might force the weights to be smaller but that can contribute to vanishing gradients.

9. Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) 1 / 1 point

☐ Vanishing gradient

☐ Exploding gradient

☒ Data augmentation

☒ Correct

☒ Dropout

☒ Correct

☐ Gradient Checking

☐ Xavier initialization

☒ L2 regularization

☒ Correct

10. Why do we normalize the inputs x ? 1 / 1 point

☐ It makes it easier to visualize the data

☐ Normalization is another word for regularization--it helps to reduce variance

☐ It makes the parameter initialization faster

☒ It makes the cost function faster to optimize

☒ Correct