Congratulations! You passed!

Grade received 80% **Latest Submission Grade** 80% **To pass** 80% or higher

Go to next item

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

- 60% train. 20% dev. 20% test
- 33% train. 33% dev. 33% test
- (selected) 98% train. 1% dev. 1% test

Expand

Correct

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 true? **1 / 1 point**

- (selected) He might be overfitting to the dev set.
- He won't be able to measure the bias of the model.
- He won't be able to measure the variance of the model.
- Not having a test set is unacceptable under any circumstance.

Expand

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. A model developed for a project is presenting high bias. One of the sponsors of the project offers some resources that might help reduce the bias. Which of the following additional resources has a better chance to help reduce the bias? **0 / 1 point**

- Use different sources to gather data and better test the model.
- Give access to more computational resources like GPUs.
- (selected) Gather more data for the project.

Expand

Incorrect
No. More data won't reduce the bias.

4. You are working on an automated check-out kiosk for a supermarket and are building a classifier for apples, bananas, and oranges. Suppose your classifier obtains a **1 / 1 point**

training set error of 19% and a dev set error of 21%. Which of the following are promising things to try to improve your classifier? (Check all that apply, suppose the human error is approximately 0%)

- [x] Use a bigger network.
- [ ] Get more training data.
- [ ] Increase the regularization parameter lambda.

Expand

Correct
Yes. This can be helpful to reduce the bias of the model, and then we can start trying to reduce the high variance if this happens.

5. What is weight decay? **1 / 1 point**

- [ ] The process of gradually decreasing the learning rate during training.
- [ ] A technique to avoid vanishing gradient by imposing a ceiling on the values of the weights.
- [x] A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration.
- [ ] Gradual corruption of the weights in the neural network if it is trained on noisy data.

Expand

Correct

6. The regularization hyperparameter must be set to zero during testing to avoid getting random results. True/False? **0 / 1 point**

- [x] True
- [ ] False

Expand

Incorrect
Incorrect. The regularization parameter affects how the weights change during training, this means during backpropagation. It has no effect during the forward propagation that is when predictions for the test are made.

7. With the inverted dropout technique, at test time: **1 / 1 point**

- [ ] You do not apply dropout (do not randomly eliminate units), but keep the 1/keep_prob factor in the calculations used in training.
- [ ] You apply dropout (randomly eliminating units) but keep the 1/keep_prob factor in the calculations used in training.
- [x] You do not apply dropout (do not randomly eliminate units) and do not keep the 1/keep_prob factor in the calculations used in training
- [ ] You apply dropout (randomly eliminating units) and do not keep the 1/keep_prob factor in the calculations used in training

Expand

Correct

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? **1 / 1 point**

- (•) Use Xavier initialization.
- ( ) Use a larger regularization parameter.
- ( ) Increase the number of cycles during the training.
- ( ) Increase the number of layers of the network.

Expand

Correct
Correct. A careful initialization can help reduce the vanishing gradient problem.

9. Which of the following actions increase the regularization of a model? (Check all that apply) **1 / 1 point**

- [ ] Increase the value of keep_prob in dropout.
- [x] Decrease the value of keep_prob in dropout.

Correct
Correct. When decreasing the keep_prob value, the probability that a node gets discarded during training is higher, thus reducing the regularization effect.

- [ ] Decrease the value of the hyperparameter lambda.
- [x] Increase the value of the hyperparameter lambda.

Correct
Correct. When increasing the hyperparameter lambda, we increase the effect of the L_2 penalization.

- [ ] Use Xavier initialization.

Expand

Correct
Great, you got all the right answers.

10. Suppose that a model uses, as one feature, the total number of kilometers walked by a person during a year, and another feature is the height of the person in meters. What is the most likely effect of normalization of the input data? **1 / 1 point**

- (•) It will make the training faster.
- ( ) It will make the data easier to visualize.
- ( ) It won't have any positive or negative effects.
- ( ) It will increase the variance of the model.

Expand

Correct

Correct. Since the difference between the ranges of the features is very different, this will likely cause the process of gradient descent to oscillate, making the optimization process longer.