Congratulations! You passed!

 $\textbf{Grade received} \ 90\% \quad \textbf{Latest Submission Grade} \ 90\% \quad \textbf{To pass} \ 80\% \ \text{or higher}$

Go to next item

1.	If you have 10,000,000 examples, how would you split the train/dev/test set?	1/1 point
	98% train. 1% dev. 1% test	
	60% train. 20% dev. 20% test	
	33% train. 33% dev. 33% test	
	∠ ⁷ 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
	He won't be able to measure the bias of the model.	
	Not having a test set is unacceptable under any circumstance.	
	He might be overfitting to the dev set.	
	He won't be able to measure the variance of the model.	
	∠ ⁷ Expand	
	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 bias, what of the following would be promising things to try? (Check all that apply.)	1 / 1 point
	Make the Neural Network deeper	
	✓ Correct	
	Get more training data	
	Increase the number of units in each hidden layer	
	✓ Correct	
	Add regularization	

	∠ [™] Expand	
	✓ CorrectGreat, you got all the right answers.	
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 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%)	1/1 point
	Use a bigger network.	
	Get more training data.	
	Increase the regularization parameter lambda.	
	∠ ⁷ 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.	Which of the following are regularization techniques? Weight decay.	1/1 point
	✓ CorrectCorrect. Weight decay is a form of regularization.	
	Gradient Checking.	
	Increase the number of layers of the network.	
	✓ Dropout.	
	 ✓ Correct Correct. Using dropout layers is a regularization technique. 	
	∠ [¬] Expand	
	✓ CorrectGreat, you got all the right answers.	
6.	The regularization hyperparameter must be set to zero during testing to avoid getting random results. True/False?	1 / 1 point
	False	
	○ True	

	Orrect Correct. 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:	0 / 1 point
	You apply dropout (randomly eliminating units) and do not keep the 1/keep_prob factor in the calculations used in training	, ,
	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 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.	
	Total apply dropout (randomly eliminating units) but keep the 1/keep_problactor in the calculations used in training.	
	∠ [¬] Expand	
	⊗ Incorrect	
8.	Increasing the parameter keep_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply)	1 / 1 point
	☐ Increasing the regularization effect	
	Reducing the regularization effect	
	✓ Correct	
	Causing the neural network to end up with a higher training set error	
	Causing the neural network to end up with a lower training set error	
	✓ Correct	
	∠ [¬] Expand	
	✓ CorrectGreat, you got all the right answers.	
9.	Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.)	1 / 1 point
	☐ Vanishing gradient	
	Data augmentation	
	✓ Correct	
	Gradient Checking	
	Xavier initialization	
	✓ L2 regularization	

∠⁷ Expand

	✓ Correct	
	✓ Dropout	
	✓ Correct	
	Exploding gradient	
	_ω [¬] Expand	
(✓ CorrectGreat, you got all the right answers.	
	uppose 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 meters. What is the most likely effect of normalization of the input data?	1 / 1 poin
	It won't have any positive or negative effects.	
	It will make the data easier to visualize.	
	It will make the training faster.	
	It will increase the variance of the model.	
	_e [⊅] 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. 	