Practical aspects of deep learning

10/10 points (100%)

Quiz, 10 questions

✓ Congratulations! You passed!	Next Item
1/1 points	
1. If you have 10,000,000 examples, how would you split the trai set?	n/dev/test
98% train . 1% dev . 1% test	
Correct	
60% train . 20% dev . 20% test	
33% train . 33% dev . 33% test	
1/1 points 2.	
The dev and test set should:	
Come from the same distribution Correct	

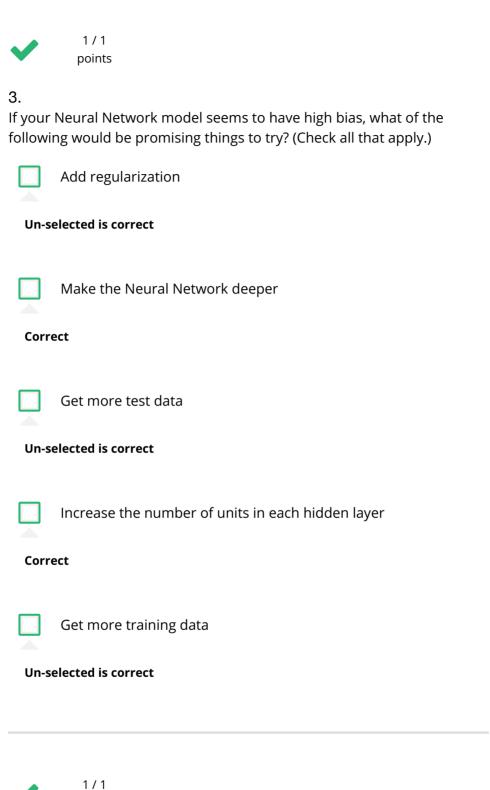
Come from different distributions

Be identical to each other (same (x,y) pairs)

Have the same number of examples Practical aspects of deep learning

10/10 points (100%)

Quiz, 10 questions



/

points

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 0.5%, and a dev set error of 7%. Which of the following are promising things to try to improve your classifier? (Check all that apply.)

Increase the regularization parameter lambda

Practical aspects of deep learning

10/10 points (100%)

Quiz, 10 questions

Un-se	Decrease the regularization parameter lambda elected is correct	
Corre	Get more training data	
Un-se	Use a bigger neural network elected is correct	
~	1/1 points	
v. Vhat is	s weight decay?	
	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.	
0	A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration.	
Corre	ect	
	A technique to avoid vanishing gradient by imposing a ceiling on the values of the weights.	

/

1/1 points

6.

What happens when you increase the regularization hyperparameter Practical aspects of deep learning

10/10 points (100%)

Ouiz.	10	questions
Quiz,		questions

0	Weights are pushed toward becoming smaller (closer to 0)			
Correct				
	Weights are pushed toward becoming bigger (further from 0)			
	Doubling lambda should roughly result in doubling the weights			
	Gradient descent taking bigger steps with each iteration (proportional to lambda)			
~	1 / 1 points			
7.				
With the inverted dropout technique, at test time:				
	You apply dropout (randomly eliminating units) but keep the 1/keep_prob factor in the calculations used in training.			
0	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			
Company				

Correct

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), but keep the 1/keep_prob factor in the calculations used in training.



1/1 points

Increasing the parameter keep_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply)

Increasing the regularization effect

Practical aspects of deep learning

10/10 points (100%)

Quiz, 10 questions

011-3	elected is correct		
Corre	Reducing the regularization effect		
	Causing the neural network to end up with a higher training set error		
Un-s	elected is correct		
	Causing the neural network to end up with a lower training set error		
Corre	ect		
	1 / 1 points of these techniques are useful for reducing variance (reducing ting)? (Check all that apply.) L2 regularization		
Corre	Dropout		
Corre	Data augmentation		
	Gradient Checking		

Un-selected is correct

Practical aspects of deep learning

10/10 points (100%)

uiz, 10 questions		Exploding gradient
	Un-s	elected is correct
		Xavier initialization
	Un-s	elected is correct
		Vanishing gradient
	Un-s	elected is correct
		1/1
		points
	10. Why do	o we normalize the inputs x ?
		It makes the parameter initialization faster
		It makes it easier to visualize the data
		Normalization is another word for regularizationIt helps to reduce variance
	0	It makes the cost function faster to optimize
	Corr	ect



