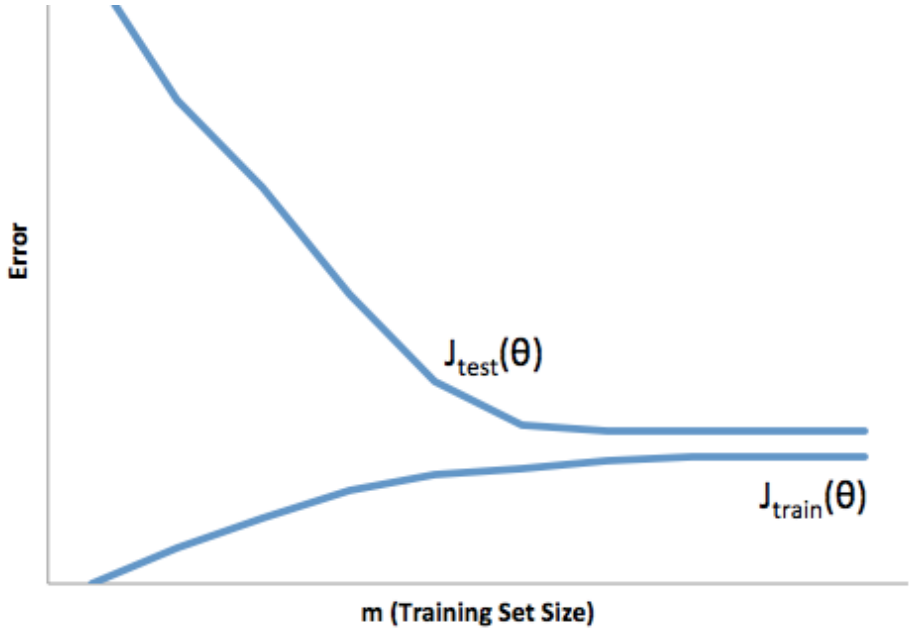
Feedback — X. Advice for Applying Machine Learning

Help Center

You submitted this quiz on **Tue 7 Apr 2015 10:41 AM CEST**. You got a score of **5.00** out of **5.00**.

Question 1

You train a learning algorithm, and find that it has unacceptably high error on the test set. You plot the learning curve, and obtain the figure below. Is the algorithm suffering from high bias, high variance, or neither?



Your Answer		Score	Explanation
○ Neither			
◉ High bias	✔	1.00	This learning curve shows high error on both the training and test sets, so the algorithm is suffering from high bias.
○ High variance			
Total		1.00 / 1.00	

Question 2

Suppose you have implemented regularized logistic regression to classify what object is in an image (i.e., to do object recognition). However, when you test your hypothesis on a new set of images, you find that it makes unacceptably large errors with its predictions on the new images. However, your hypothesis performs **well** (has low error) on the training set. Which of the following are promising steps to take? Check all that apply.

Your Answer		Score	Explanation
☑ Get more training examples.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Adding more training data will increase the complexity of the training set and help with the variance problem.
☐ Try decreasing the regularization parameter λ.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Decreasing the regularization parameter will increase the overfitting, not decrease it.
☐ Use fewer training examples.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Using fewer training examples will only exacerbate the overfitting.
☑ Try increasing the regularization parameter λ.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Increasing the regularization parameter will reduce overfitting and help with the variance problem.
Total		1.00 / 1.00	

Question 3

Suppose you have implemented regularized logistic regression to predict what items customers will purchase on a web shopping site. However, when you test your hypothesis on a new set of customers, you find that it makes unacceptably large errors in its predictions. Furthermore, the hypothesis performs **poorly** on the training set. Which of the following might be promising steps to take? Check all that apply.

Your Answer		Score	Explanation
☐ Try increasing the regularization parameter λ.	✔	0.25	The poor performance on both the training and test sets suggests a high bias problem. Increasing regularization will decrease the fit of the hypothesis to the data, exacerbating the high bias problem.
☐ Use fewer training examples.	✔	0.25	Using fewer training examples should never improve test set performance, as the model has fewer data points from which to learn.
☑ Try decreasing the regularization parameter λ.	✔	0.25	The poor performance on both the training and test sets suggests a high bias problem. Decreasing the regularization parameter will allow the hypothesis to fit the data more closely, improving both training and test set performance.
☑ Try to obtain and use additional features.	✔	0.25	The poor performance on both the training and test sets suggests a high bias problem. Using additional features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.
Total		1.00 / 1.00	

Question 4

Which of the following statements are true? Check all that apply.

Your Answer		Score	Explanation
☑ Suppose you are training a regularized linear regression model. The recommended way to choose what value of regularization parameter λ to use is to choose the value of λ which gives the lowest **cross validation** error.	✔	0.25	The cross validation lets us find the "just right" setting of the regularization parameter given the fixed model parameters learned from the training set.
☑ The performance of a learning algorithm on the	✔	0.25	The learning algorithm finds parameters to minimize training set error, so the performance should be better on the training set than the test set.

training set will typically be better than its performance on the test set.			
☐ Suppose you are training a regularized linear regression model.The recommended way to choose what value of regularization parameter λ to use is to choose the value of λ which gives the lowest **training set** error.	✔	0.25	You should not use training error to choose the regularization parameter, as you can always improve training error by using less regularization (a smaller value of λ). But too small of a value will not generalize well on the test set.
☐ It is okay to use data from the test set to choose the regularization parameter λ, but not the model parameters (θ).	✔	0.25	You should not use test set data in choosing the regularization parameter, as it means the test error will not be a good estimate of generalization error.
Total		1.00 / 1.00	

Question 5

Which of the following statements are true? Check all that apply.

Your Answer		Score	Explanation
☑ If a learning algorithm is suffering from high bias, only adding more training examples may **not** improve the test error significantly.	✔	0.25	With high bias, the model is not fitting the training data currently present, so adding more data is unlikely to help.
☑ A model with more parameters is more prone to overfitting and typically has higher variance.	✔	0.25	More model parameters increases the model's complexity, so it can more tightly fit data in training, increasing the chances of overfitting.
☑ When debugging learning	✔	0.25	The shape of a learning curve is a good indicator

algorithms, it is useful to plot a learning curve to understand if there is a high bias or high variance problem.			of bias or variance problems with your learning algorithm.
☐ If the training and test errors are about the same, adding more features will **not** help improve the results.	✔	0.25	If the two errors are the same, then the model has high bias, so adding more features will be helpful.
Total		1.00 / 1.00	