Revision

High variance

- indicated by gap in errors between training and testing data sets.
- Algorithm has overfit the data for the training data set.
- Increasing the regularization parameter will reduce overfitting

The recommended way to choose a value of regularization parameter λ to use is to choose the lowest cross validation error. You should not use the training data set for this purpose.

1 Week 6: Advice for Applying Machine Learning

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?

(i) Neither

(ii) High variance

(iii) High bias

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.

(i) Try increasing the regularization parameter λ.

(ii) Try evaluating the hypothesis on a cross validation set rather than the test set.

(iii) Try using a smaller set of features.

(iv) Try decreasing the regularization parameter λ.

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.

(i) Try decreasing the regularization parameter λ.

(ii) Use fewer training examples.

(iii) Try evaluating the hypothesis on a cross validation set rather than the test set.

(iv) Try adding polynomial features.

Question 4.

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

(i) 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 test set error.

(ii) 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.

(iii) The performance of a learning algorithm on the training set will typically be better than its performance on the test set.

(iv) 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.

Question 5.

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

(i) A model with more parameters is more prone to overfitting and typically has higher variance. SELECTED

(ii) If the training and test errors are about the same, adding more features will not help improve the results. SELECTED

(iii) If a learning algorithm is suffering from high variance, adding more training examples is likely to improve the test error.

(iv) If a learning algorithm is suffering from high bias, only adding more training examples may not improve the test error significantly.

Week 6: Machine Learning System Design

Question 1

You are working on a spam classification system using regularized logistic regression. "*Spam*" is a positive class (y = 1) and "*not spam*" is the negative class (y = 0). You have trained your classifier and there are m = 1000 examples in the cross-validation set. The chart of predicted class vs. actual class is:

	Actual Class: 1	Actual Class: 0
Predicted Class: 1	85	890
Predicted Class: 0	15	10

For reference:

- Accuracy = (true positives + true negatives) / (total examples)
- Precision = (true positives) / (true positives + false positives)
- Recall = (true positives) / (true positives + false negatives)
- F1 score = (2 × precision × recall) / (precision + recall)

What is the classifier's precision (as a value from 0 to 1)?

Enter your answer in the box below. If necessary, provide at least two values after the decimal point.

0.09

1.1 Question 1

CORRECT Suppose a massive dataset is available for training a learning algorithm. Training on a lot of data is likely to give good performance when two of the following conditions hold true. Which are the two?

(i) WRONG When we are willing to include high order polynomial features of x (such as x_1^2, x_2^2, x_1x_2, etc.).

(ii) SELECTED The features x contain sufficient information to predict y accurately. (For example, one way to verify this is if a human expert on the domain can confidently predict y when given only x).

(iii) SELECTED We train a learning algorithm with a large number of parameters (that is able to learn/represent fairly complex functions).

(iv) We train a learning algorithm with a small number of parameters (that is thus unlikely to overfit).

1.2 Question 3.

Suppose you have trained a logistic regression classifier which is outputing $h\theta(x)$. Currently, you predict 1 if $h\theta(x)$ $\geq$threshold, and predict 0 if $h\theta(x)$ ¡ threshold, where currently the threshold is set to 0.5. Suppose you decrease the threshold to 0.1. Which of the following are true? Check all that apply.

(i) The classifier is likely to have unchanged precision and recall, but lower accuracy.

(ii) The classifier is likely to now have higher precision.

(iii) SELECTED The classifier is likely to now have higher recall.

(iv) The classifier is likely to have unchanged precision and recall, but higher accuracy.

(v) The classifier is likely to have unchanged precision and recall, and thus the same F1 score.

(vi) The classifier is likely to have unchanged precision and recall, but higher accuracy.

(vii) SELECTED The classifier is likely to now have lower precision.

(viii) The classifier is likely to now have lower recall.

Question 4.

WRONG Suppose you are working on a spam classifier, where spam emails are positive examples (y=1) and non-spam emails are negative examples (y=0). You have a training set of emails in which 99% of the emails are non-spam and the other 1% is spam. Which of the following statements are true? Check all that apply.

- WRONG If you always predict spam (output y=1), your classifier will have a recall of 0% and precision of 99%.
- WRONG If you always predict non-spam (output y=0), your classifier will have a recall of 0%.
- CORRECT If you always predict spam (output y=1), your classifier will have a recall of 100% and precision of 1%.
- CORRECT If you always predict non-spam (output y=0), your classifier will have an accuracy of 99%.

	$predictT$	$predictF$
$actualT$		
$actualF$		

Question 5.

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

(i) CORRECT Using a very large training set makes it unlikely for model to overfit the training data.

(ii) WRONG If your model is underfitting the training set, then obtaining more data is likely to help.

(iii) CORRECT The "error analysis" process of manually examining the examples which your algorithm got wrong can help suggest what are good steps to take (e.g., developing new features) to improve your algorithm's performance.

(iv) WRONG It is a good idea to spend a lot of time collecting a large amount of data before building your first version of a learning algorithm.

(v) WRONG After training a logistic regression classifier, you must use 0.5 as your threshold for predicting whether an example is positive or negative.