

Name:

Student ID#:

Statistical Pattern Recognition (CE-725)
Department of Computer Engineering
Quiz #4 (Classification – Introduction) - Spring 2012

1. (20 points) For each of the following statements, write “True” if it is mostly true, or “False” if it is mostly false. Then, explain why or how the statement is true or false:

a) Cross-validation can reveal overfitting.

Sol:

True. If accuracy on the training folds is higher than on the test folds, then overfitting is revealed.

b) Overfitting is a danger when learning a classifier, but not when doing unsupervised learning.

Sol:

False. Informally, overfitting means finding patterns in the data that are spurious because they are true in the training data but false in the test data. It is certainly possible to find spurious patterns in unlabeled data.

2. (30 points) You are reviewer for the international Conference on Algorithms and you read papers with the following experimental setups. Would you accept or reject each paper? Provide a one sentence justification.

a) “My algorithm is better than yours. Look at the test error rates! (Footnote: reported results for parameter $L=1.7894898783$.)”

Sol:

Reject A L with 10 decimal places suggests a highly tuned solution, probably looking at the test data.

b) “My algorithm is better than yours. Look at the test error rates! (Footnote: reported results for best value of L.)”

Sol:

Reject Choosing L based on the test data?

c) “ My algorithm is better than yours. Look at the test error rates! (Footnote: reported results for best value of L, Chosen with 10-fold cross validation.)”

Sol:

Accept- Cross validation is the appropriate method for selecting parameters.