

Name:

Student ID#:

Statistical Pattern Recognition (CE-725)
Department of Computer Engineering
Quiz #1 solution (Overview & Introductory Materials)
Spring 2010

1.a (6 points) What is the problem of under-fitting and over-fitting in pattern classification?

Under-fitting: Inability to successfully classify unseen instances as well as seen instances. A classifier that is not sufficiently complex may lead to under-fitting.

Over-fitting: Excellent classification in seen instances (once they have been seen) but, poor classification in unseen instances. A classifier that is too complex may fit the noise, not just the samples, and may lead to over-fitting.

1.b (4 points) Does the size of the data set affect the over-fitting problem on a given model? Why?

For a given model, the over-fitting problem become less severe as the size of the data set increases. The following figure illustrates this fact. A rough heuristic is that the number of data points should be no less than some multiple (say 5 or 10) of the number of adaptive parameters in the model.

