

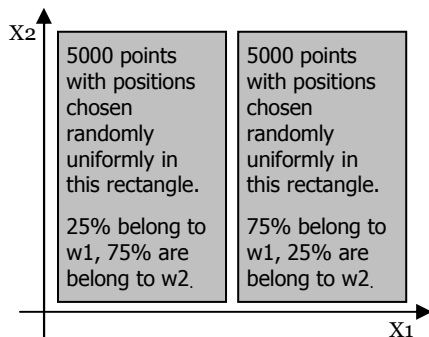
Name:

Student ID#:

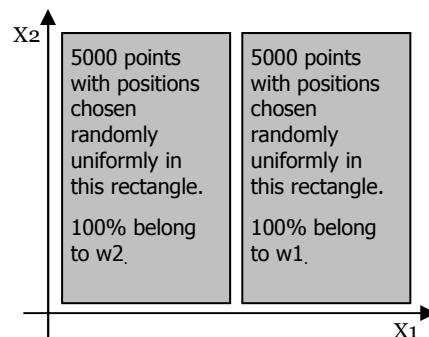
Statistical Pattern Recognition (CE-725)
Department of Computer Engineering
Quiz #7 Solution (Non-Parametric Modeling) - Spring 2011

Consider the following recipes for creating training and test set for a problem. As you see, the training set is noisy but the test set is not.

Recipe for making training set of 10,000 data points with two real valued inputs and one binary output class:



Recipe for making test set of 10,000 data points with two real valued inputs and one binary output class:



- a. (2.5 pts) What is the expected leave-one-out cross-validation error on the training set using one-nearest-neighbor?

$$\text{error} = 1/2 * [(3/4 * 1/4) + (1/4 * 3/4)] + 1/2 * [(1/4 * 3/4) + (3/4 * 1/4)] = 3/8$$

- b. (2.5 pts) What is the expected test set error if we train on the training set, test using the test set, and use one-nearest-neighbor?

$$\text{error} = 1/2 * (1 * 1/4) + 1/2 * (1 * 1/4) = 2/8$$

- c. (2.5 pts) What is the (approximate) expected leave-one-out cross-validation error on the training set using 21-nearest-neighbors?

In average $0.75 * 21=15$ neighbors of each sample point will belong to one of classes. Then the error will be:

$$\text{error} = 1/2 * 1/4 + 1/2 * 1/4 = 1/4$$

- d. (2.5 pts) What is the (approximate) expected test set error if we train on the training set, test using the test set, and use 21-nearest-neighbor?

$$\text{error} = 1/2 * 0 + 1/2 * 0 = 0 \text{ (same argument as above)}$$