Error Rates

- ▶ We can evaluate error rates by means of a training sample (to construct the discrimination rule) and a test sample.
- ▶ An optimistic error rate is obtained by reclassifying the training data. (In the **training data** sets, how many cases were misclassified). This is known as the **apparent error rate**.
- ▶ The apparent error rate is obtained by using in the training set to estimate the error rates.
- ▶ It can be severely optimistically biased, particularly for complex classifiers, and in the presence of over-fitted models.

- If an independent test sample is used for classifying, we arrive at the **true error rate**.
- The true error rate (or conditional error rate) of a classifier is the expected probability of misclassifying a randomly selected pattern.
- It is the error rate of an infinitely large test set drawn from the same distribution as the training data.

Misclassification Cost

- ▶ As in all statistical procedures it is helpful to use diagnostic procedures to asses the efficacy of the discriminant analysis. We use **cross-validation** to assess the classification probability. Typically you are going to have some prior rule as to what is an **acceptable misclassification rate**.

- ▶ Those rules might involve things like, "what is the cost of misclassification?" Consider a medical study where you might be able to diagnose cancer.

- ▶ There are really two alternative costs. The cost of misclassifying someone as having cancer when they don't. This could cause a certain amount of emotional grief. Additionally there would be the substantial cost of unnecessary treatment.

There is also the alternative cost of misclassifying someone as not having cancer when in fact they do have it.
A good classification procedure should

- ▶ result in few misclassifications
- ▶ take **prior probabilities of occurrence** into account
- ▶ consider the cost of misclassification

For example, suppose there tend to be more financially sound firms than bankrupt firm. If we really believe that the prior probability of a financially distressed and ultimately bankrupted firm is very small, then one should classify a randomly selected firm as non-bankrupt unless the data overwhelmingly favor bankruptcy.

- There are two costs associated with discriminant analysis classification: The true misclassification cost per class, and the expected misclassification cost (ECM) per observation.
- Suppose there we have a binary classification system, with two classes: class 1 and class 2.
- Suppose that classifying a class 1 object as belonging to class 2 represents a more serious error than classifying a class 2 object as belonging to class 1.
- There would an assignable cost to each error. $c(i|j)$ is the cost of classifying an observation into class j if its true class is i.

Misclassification

The costs of misclassification can be defined by a cost matrix.

	Predicted Class 1	Predicted Class 2
Class 1	0	$c(2|1)$
Class 2	$c(1|2)$	0

Expected cost of misclassification (ECM)

- ▶ Let p_1 and p_2 be the prior probability of class 1 and class 2 respectively. Necessarily $p_1 + p_2 = 1$.
- ▶ The conditional probability of classifying an object as class 1 when it is in fact from class 2 is denoted $p(1|2)$.
- ▶ Similarly the conditional probability of classifying an object as class 2 when it is in fact from class 1 is denoted $p(2|1)$.

$$ECM = c(2|1)p(2|1)p_1 + c(1|2)p(1|2)p_2$$

(In other words: the sum of the cost of misclassification times the (joint) probability of that misclassification.

- ▶ A reasonable classification rule should have ECM as small as possible.