STAT 802 Multivariate Analysis October 28, 2003

Ch 10: Canonical Correlations

We will leave until later (if time permits):

Basic idea is to determine the relationship between two sets of variables:

Not Multivariate Regression because no dependent variable. Also, may be more than one relationship. Two linear combos with highest correlation, then look in perpendicular directions in both set-spaces for the next pair of linear combos with highest correlation.

Examples mentioned in the intro to Ch 10:

Set I: arithmetic speed, arithmetic power
Set II: reading speed, reading power

Set I: govt policy vbles
Set II: economic goal vbles

Set I: college performance vbles
Set II: precollege achievement vbles

Useful in principle, but not much used in practice.

Ch 11: Discrimination and Classification:

Suppose you have some multivariate data for each of two or more groups of items:

Discrimination: How do you describe the differences among the groups in terms of the variables?

Classification: How do you decide the group membership of items based only on the variables?

Sometimes an index can be employed to do both jobs (or k-1 indices for k groups).

The 2-D picture is instructive for thinking about these procedures: See p 585.

What characteristic of the data determines the dot type?
Why is a line used to separate the groups?
Why are some points on the wrong side of the discriminant line?
Is the line related to an “index”?
How is the index used to classify items of unknown group membership?
Are all points given the same weight in the discrimination problem?

How do you summarize the degree of misclassification that the discriminant line implies?
Is the rarity of the groups in the population a factor in determining misclassification error?
Is the cost of misclassification the same for all points? Is this important?

An even simpler picture of the 1-D discrimination/classification problem is also useful. See Fig 11.3 on p 586.
For any given variate value, two likelihoods can be computed. Does it make sense to assign that value to the group whose likelihood is the greatest?
What does this imply if the costs and priors are equal for the two groups? Summary p 588.

Slightly more complex situation – two multivariate normal populations – density ratios determined by difference of exponents (of normal densities). Is (11-13) useful?
Some help to write it as $(\mu_1 - \mu_2)'\Sigma^{-1}(x - \bar{\mu})$. Explain in terms of projections?

Equivalent sample statistics version p 592. But note pooling of Σ estimates – valid only if covariance of both groups is the same (or correlations if using standardized data).
Effect of keeping distinct $\Sigma_1 \neq \Sigma_2$ is that index is quadratic – see (11-25) p 597.

Misclassification probabilities can be computed explicitly for two normal populations. (11-27)

Note difference between apperent misclassification rate (APER) and the actual error rate (AER). APER will usually be smaller than AER since the same data is used to estimate the discriminant function as to test this discriminant function. (p 602) Holdout method gives a way of avoiding this bias (p 602-3)

Exercise: Estimate the (AER) for the salmon data (p 607). Use the quadratic discriminant function. Use Fig 11.7 to suggest an heuristic explanation why unequal covariances require a nonlinear discriminant function.

Note the seemingly different approach of the Fisher Linear Discriminant Function which produces the same linear discriminator. (pp 609-612)

Extension to several populations.

Presentation topics????