Last time we were working our way through the Discrimination and Classification (Ch 11).

We were discussing the logic of the "maximum likelihood" approach for the discrimination problem in the simple case of equal priors and equal costs of misclassification. The density ratio being > or < 1 at x is the same as choosing the max likelihood population at x.   See p 588.   For cases of two equi-variance normal distributions, this max likelihood rule produces a decision threshold at the point where the densities are equal. This is the threshold that minimizes the total misclassification rate
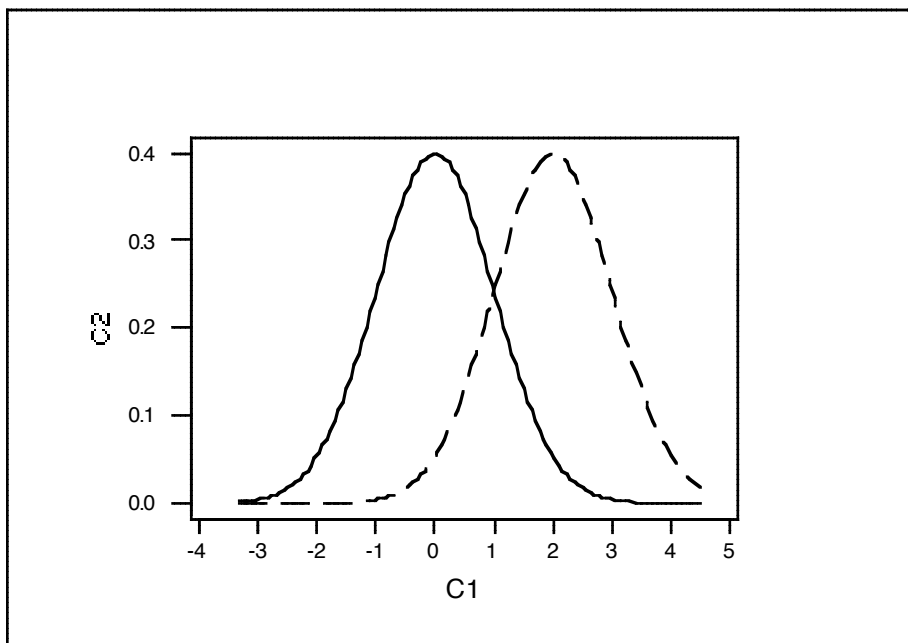
It is easy to see from the graph why the threshold at the intersection (the ml threshold) has to be the minimum classification error threshold – just consider what happens to the two error areas as you move away from the intersection. The increase in one error is greater than the decrease in the other one.

However, this tidy situation does not hold for cases like the ones shown on p 598. Instead of a single threshold associated with two contiguous regions (one for pop 1 and the other for pop 2), you get one of the regions split. The maximum lielihood criterion is still reasonable, but it leads to a more complicated decision map.

When we move to two (or more) dimensions (2-variate or p-variate data) , the boundary of the decision regions is more complicated. However, for two equal-covariance normal populations, the boundary will be a straight line. (See its equation in x_0 on p 592) This is an estimate but a similar property would hold for population values). The straight line boundary for the density equality can be deduced from the symmetry – the locus of points

equidistant from the two centroids. With this insight, we can see that unequal covariances will cause the boundary to be curved – in the normal case it is quadratic. See formula p 597 (which is more general since it applies when costs and priors may not have ratio 1.)

Note that it is often easier to specify the relative costs or relative priors than it is to specify the actual costs and priors. And it is the ratios that are needed in the classification procedure.

See Notes from Oct 28 – we did not discuss the last part ….