Form of Discriminant vs. Manner of Estimation

LDA and logistic regression use the same *form* of linear discriminant, and the same form for class probabilities. But they *estimate* the parameters of this form differently.

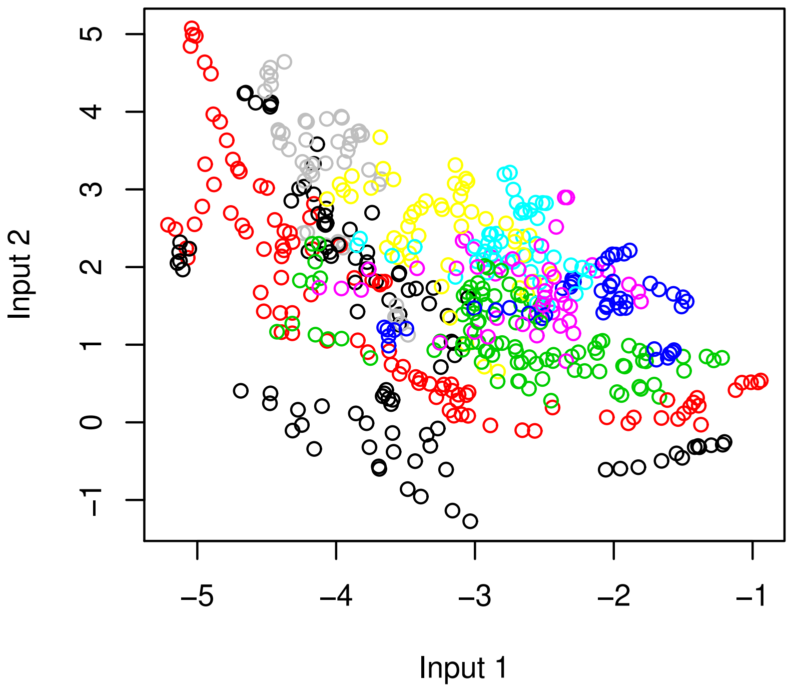
- LDA uses estimates from a *joint* model of inputs and class, assuming Gaussian class densities with a common covariance.
- Logistic regression uses estimates from a *conditional* model of class given inputs.
- A compromise mentioned in the book: Use the LDA estimates except for the intercept, which is set to maximize classification accuracy on the training data.

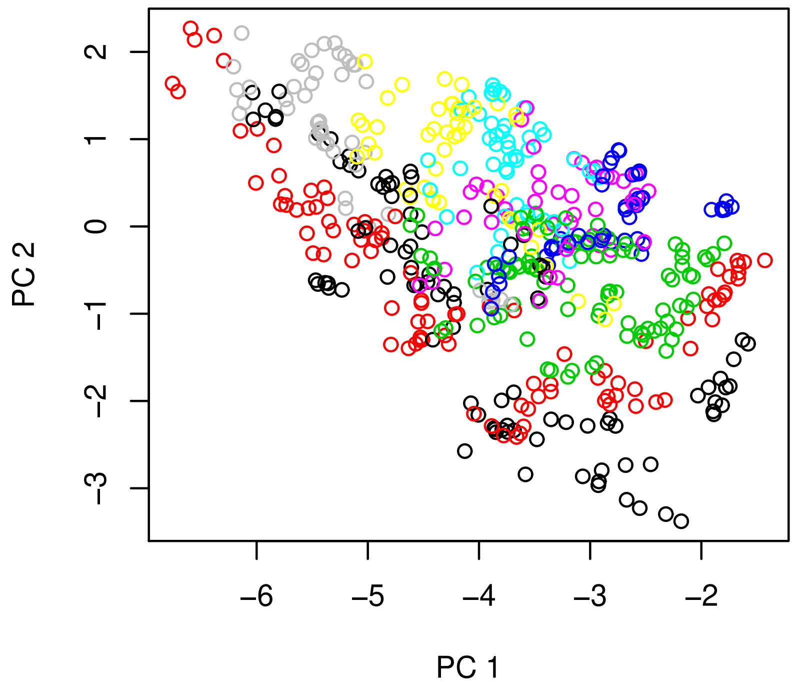
LDA should be better if the assumption of Gaussian class densities with the same covariance is true. Otherwise, it seems wiser to use logistic regression.

Example: Vowel Recognition Problem

As a small component of computer a speech recognition system, we can look at data on the speech waveforms of spoken vowels. The task is to recognize the vowels (11 classes) from the waveform data (10 inputs). There are 528 training cases, and 462 test cases. (See page 85 in the text.)

Here is a plot of the training cases, using the first two inputs, and the first two principal components:





Results on the Vowel Recognition Problem

Here are error rates on the training and test sets when using LDA and logistic regression (LR), with all inputs, with the first two inputs, and with two or four principal components. Results using QDA are also shown.

	train	test
LDA, all inputs	0.316	0.556
LDA, 2 inputs	0.527	0.481
LDA, 2 PCs	0.521	0.498
LDA, 4 PCs	0.477	0.483
QDA, all inputs	0.01	0.53
LR, all inputs	0.222	0.513
LR, 2 inputs	0.509	0.481
LR, 2 PCs	0.491	0.468
LR, 4 PCs	0.426	0.463

(See also page 85 in the text; results not there are mine.)