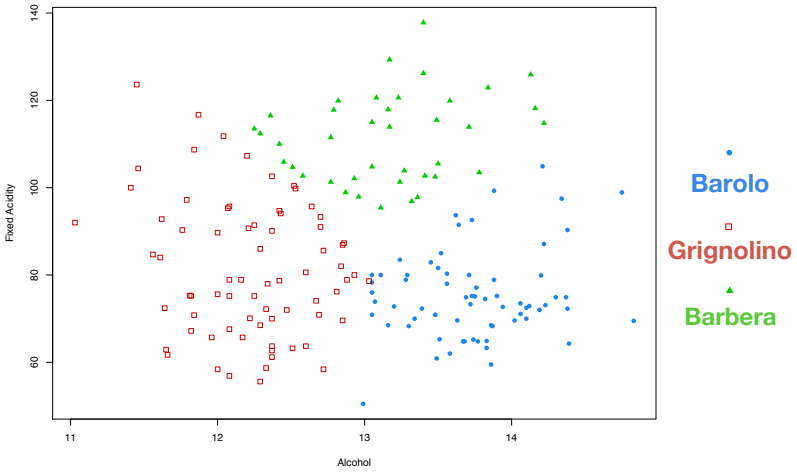
Programming Exercise – ML Course
Linear Models for Classification



1. Load the data available on LearnIT that consists in bidimensional features of some wine samples (the x.txt file), and labels for binary/ternary classification (the y_bin.txt and y_1ofK.txt files). You can use for example the numpy.loadtxt function. Give a quick look at the data.

2. Plot a scatter plot of the data. We will give more details about this data set during Monday's lecture.

3. Create a training set by randomly selecting 80% of the observations. Keep the remaining 20% as a test set. Be careful: you need to select the same observations for both the design matrix and the labels.

4. For the binary classification problem (Barolo vs rest), build a classifier using the Gaussian generative model described by Bishop in Sections 4.2.1 and 4.2.2. That model will be the main focus on Monday's lecture, and is also the theme of this week's maths exercises (4.8, 4.9, and 4.10). Compute the training classification error and the test error: what's the largest?

5. Same question for the ternary classification problem.