Department of Computer Science

California State University, Channel Islands

MATHCOMPPH-546: Pattern Recognition

Lesson 7 phys546 Discriminant_xls file 7B

Student Name: Sandipta Subir Khare

Student Major: Computer Science

1. **Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA) have different purposes. The former operates on the total (unlabeled) dataset to find the directions that contain the maximum variance; the latter operates on labeled data to find the directions which are best at distinguishing the labeled classes. In general, the results (principal components and canonicals, respectively) will be quite different. However, for special examples of data the principal components and canonicals could be in the same directions. (1) Draw an example of two-class, two-dimensional data such that PCA and LDA find the same directions. (2) Draw an example of two-class, two-dimensional data such that PCA and LDA find totally different directions.**

If we put specificity and sensitivity 90% and prevalence as 5% then 50 people have the disease and 950 people don't have the disease. With the sensitivity and specificity of 90%, the test will correctly pick up 90% or 45 people with the disease and 90% or 855 individuals without the disease. The remaining 5 individuals would falsely classify as negatives and 95 individuals as false positives. So, overall, we get 140 people who test positive.

So, PPV = 45/140 * 100 = 32%

2. **The data in socioeconomic.jmp consists of five socioeconomic variables/ features for 12 census tracts in the LA Metropolitan area. (a) Use the Multivariate platform to produce a scatterplot matrix of all five features. (b) Conduct a principal component analysis (on the correlations) of all five features. Considering the eigenvectors, which are the most useful features? Considering the eigenvalues, how many principal components would you use in subsequent analysis?**

We can do this by creating variable named targets with class labels named targets is a 1-by-N vector containing the class labels and, in this case, 0, 1 and 2. We can save these as a mat file. We can open Classifier, read data when d>2, a feature selection GUI will open and request the user to preprocess the data such as by PCA to yield 2-D data, compatible with the display and use appropriate classifiers.