Linear Data Analysis
Patterns - Linear Discriminant Analysis

Cain Susko

Queen's University
School of Computing

March 16, 2022

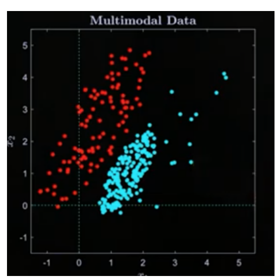
a Finding Patterns in Labeled Data

This section will focus on how to adapt PCA to analyze data that has labels. The main conectps are what does PCA do when we expolre is on the means of the labeled data; how does PCA and the Rayliegh Quotient help us sove this problem; what is LDA - optimization.

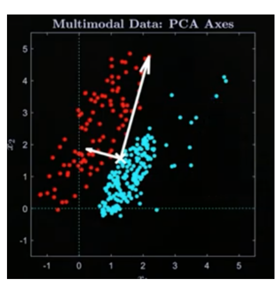
What can PCA Do? PCA is a way of explaining, not predicting data. PCA scores are used to reduce the dimensionality of a dataset. Clustering will often work better after PCA.

b Example of PCA for Labeled Data

given the data generated by prof. Ellis: 300 points in data matrix A. the datapoints are inside 3 elipses, where 2 are merged. The labels used are $y_j \in \{-1, +1\}$. We must find the mean, scatter matrix, and PCA of the data and. Additionally, we must Carry the labels from the data onto the loading vectors and designate a colour to each label for visual assesment. The labeled, coloured data is the following:

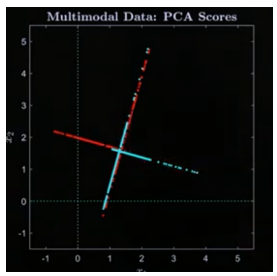


After PCA, the data look like so:



where the x inidcates the mean of all of data, the large arrow indicates the first loading vector and the smaller arrow indicated the second loading vector. Thus, we can see that the loading vectors are aligned with the major and minor axis of the original data.

Projection When we then project the labels onto these axis (the 2 loading vectors) we can see that the second loading vector may actually be doing a better job than the first loading vector at distinguishing the labels.



PCA on the Mean If we indicate the mean of the red and cyan clusters (as x's in their respective colours) and then perform PCA on the means of the data we get the magenta axis which is directed between the 2 means and goes through the means of the overall data.

Data With Labels: So Far

- we can colour data labels using the `gsatter` function.
- doing PCA with a scatter matrix is finding the maximum variance which is the first loading vector from the PCA.
- PCA treats all data uniformly

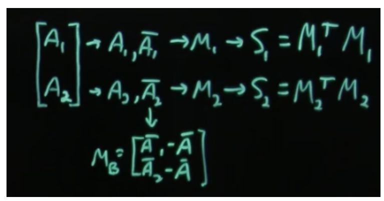
c Fishers Linear Discriminant

there are many types of scatter matricies:

S_t Total scatter $= M^\top M$

S_w Scatter within labels $= S_1 + S_2$

S_b Scatter between labels $= M_b^\top M_b$



We thus have 2 objectives:

$$\vec{v}_{max} \in S_B = \max_{\vec{u} \in \mathbb{R}^n} R(S_B)$$

$$\vec{v}_{min} \in S_w = \max_{\vec{u} \in \mathbb{R}^n} R(S_w^{-1})$$

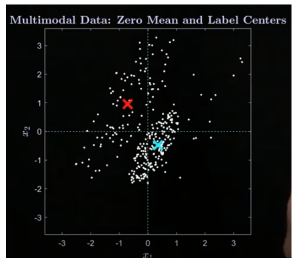
Thus, Fishers Linear Discriminant (LDA) is:

$$\vec{v} = \max_{\vec{u} \in \mathbb{R}^n} S_w^{-1} S_B$$

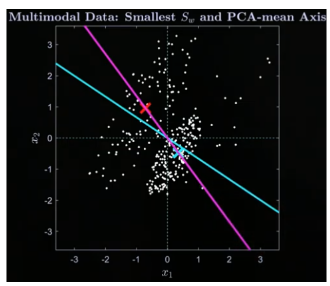
therefore, the result (maximum eigenvector within $S_w^{-1} S_B$) is Fishers Linear Discriminant that simultaneously maximizes the between label scatter and minimizes the within label scatter.

d Example of LDA for Labeled Data

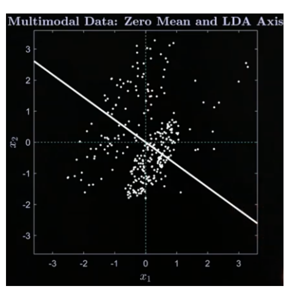
given the data with 2 labels and their means (cyan and red x's)



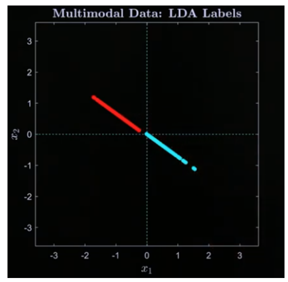
after finding the minimum PCA Axis (cyan, the axis that minimizes the within label scatter) and the maximum PCA axis (magenta, the axis that maximizes the between label scatter), which goes through the 2 means of the data.



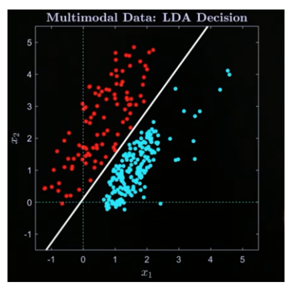
When we use Linear Discriminant Analysis, we get the white line, which simultaneously maximizes the between label scatter and minimizes the within label scatter.



when we then project the labels to the LDA axis we can see that the line seperates the 2 labels very well:



we can also derive a seperating hyper plane from the LDA Axis by choosing a point along the Axis that maximizes the accuracy of the seperation for the 2 labels within the data.



The hyper-plane will be orthogonal to the LDA axis.

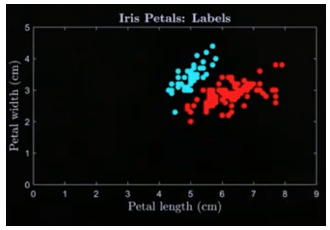
d Example of LDA for Iris Data

Recall The Data is made up of 2 measurments: petal length and width. The labels we will use are:

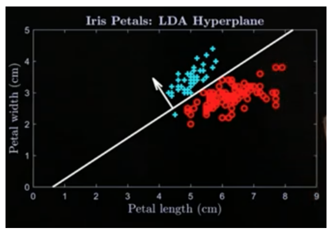
B Beach-Head plant

P Purple plant

when we plot these with width as the y axis and length as the x axis:



After applying LDA to the iris data we get the LDA axis as the arrou and the hyper-plane of seperation is the best seperation of the 2 labels orthogonal to the LDA axis.



while there is still an anomalous plants, the hyper plane is a much better way of seperating the 2 labels than say, k-means clustering.

LDA: HyperPlane From Axis

1. Project onto axis $\vec{w} = \vec{v}_{max}(S_w^{-1} S_B)$
2. recall: hyperplane needs bias scalar b
3. Thus: there will be further optimization with b, possibly using an ROC curve (covered in the next lesson)

Learning Summary

Students should now be able to:

- compute matricies from labeled data
- compute fishers discriminant
- project labels onto LDA axis to verify if LDA is working as intended