LINEAR DISCRIMINANT ANALYSIS

1

Agenda

2

- Recap of PCA
  - Major Limitations
- Dimensionality Reduction in Classification
  - Key Requirements
- Let's see how PCA fails!
- Linear Discriminant Analysis (LDA)
- PCA versus LDA

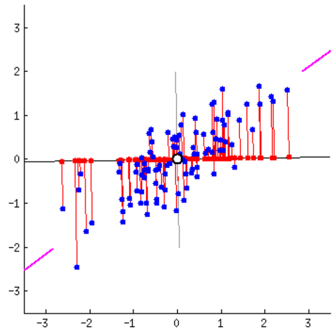
LDA

2

PCA: A Quick Recap

3

- A dimensionality reduction technique
  - Finds the most accurate lower dimensional representation based on maximum variance directions



LDA

3

PCA: A Quick Recap

4

- Key Features
  - Unsupervised (does not utilize labels)
  - Focuses on variance (not always useful or relevant)

LDA

4

PCA: Limitations

5

- Limitations
- Not suitable for discrimination/classification
  - Does not take class information into consideration
  - PCA is an optimal dimensionality-reduction technique for data representation, not for discrimination/classification problems

LDA

5

Dimensionality Reduction in Classification

6

- A simple real-world classification example: Identifying species of a fish on a conveyor belt
  - Species: Sea bass and salmon

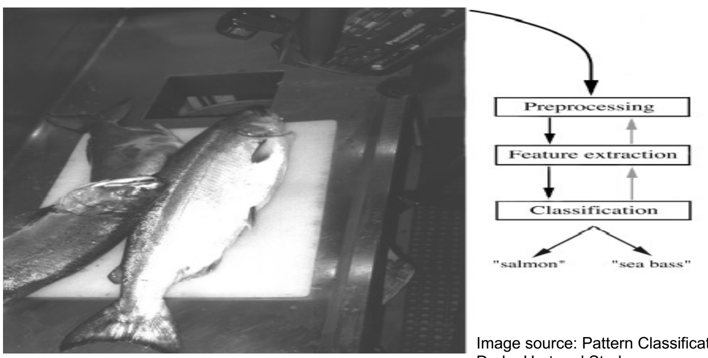


Image source: Pattern Classification by Duda, Hart and Stork

6

Dimensionality Reduction in Classification

7

- Single feature-based classification
  - Feature: Length

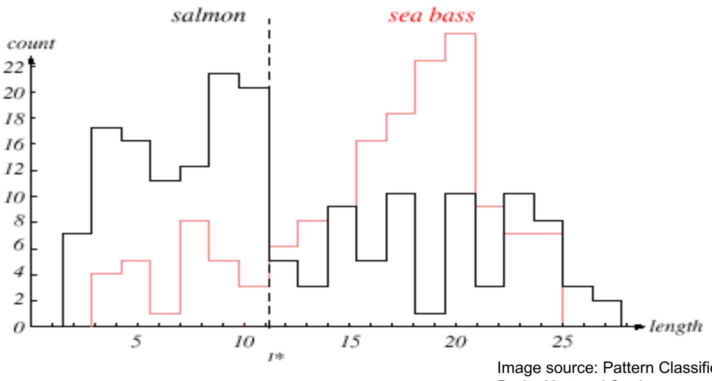


Image source: Pattern Classification by Duda, Hart and Stork

7

Dimensionality Reduction in Classification

8

- Single feature-based classification
  - Feature: Lightness

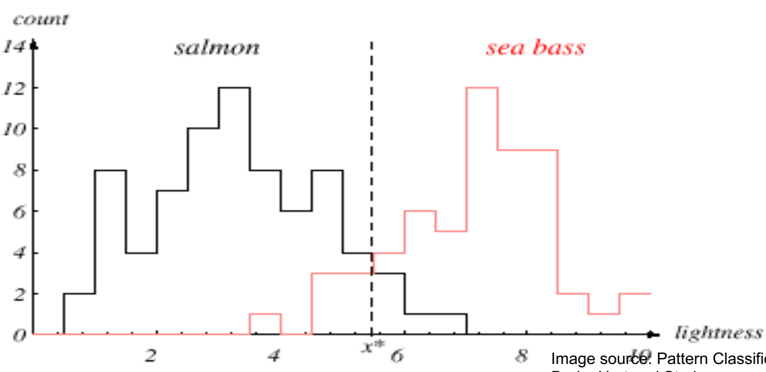


Image source: Pattern Classification by Duda, Hart and Stork

8

Dimensionality Reduction in Classification

- Two features for classification

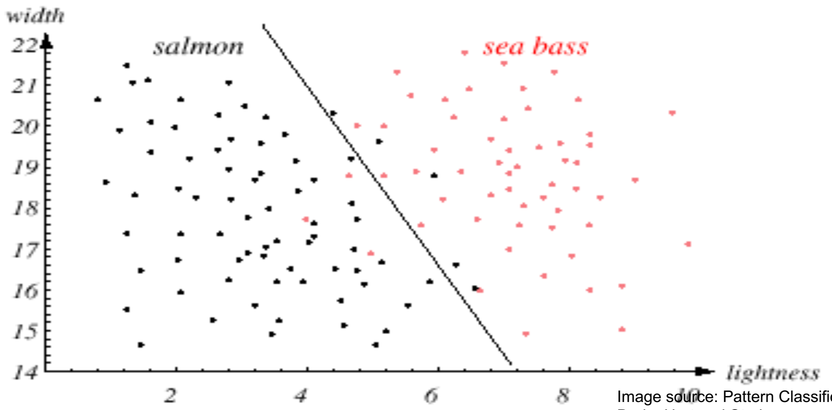
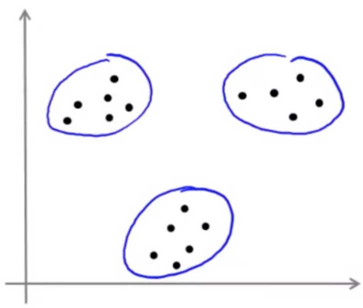


Image source: Pattern Classification by Duda, Hart and Stork

9

Dimensionality Reduction in Classification

- Key Requirements for Classification
  - Each class is as tight (compact) as possible
  - Their centroids are as far from each other as possible

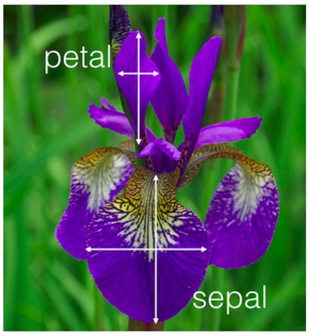


Ideal Distribution

10

Let's see how PCA Fails!

- PCA is not an effective dimensionality technique for classification
- The Iris Dataset

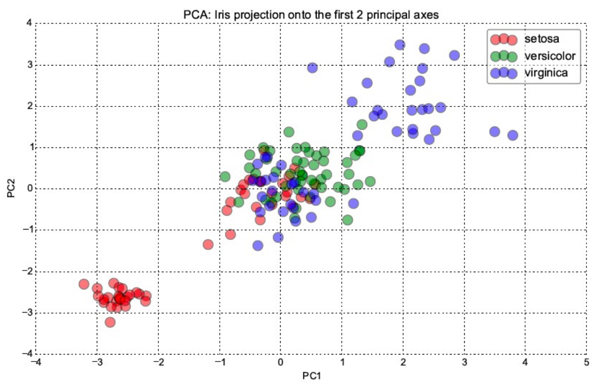


	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

One of the earliest known datasets used for evaluating classification methods, widely used in ML

Let's see how PCA Fails!

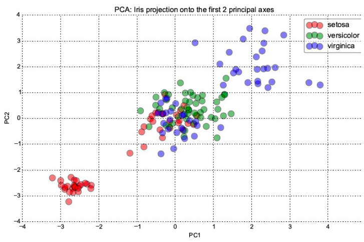
- Problem: Given a new measurement of these features, predict the Iris species based on a projection onto a low-dimensional space

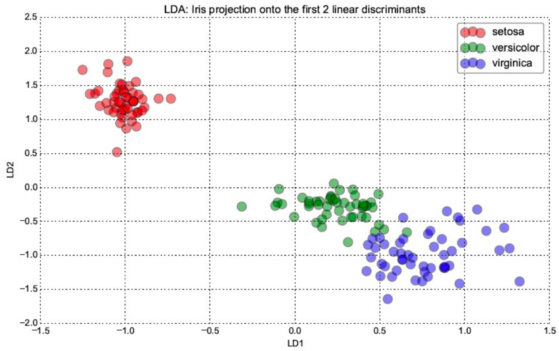


Let's see how PCA Fails!

13

- **Conclusion:** PCA may not be ideal to separate the classes well

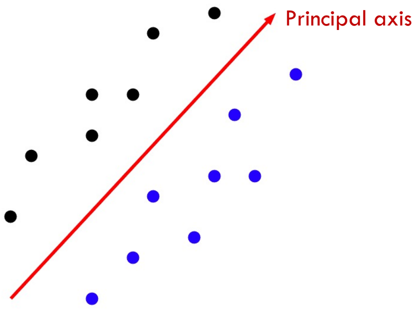




13

Why PCA Fails

14



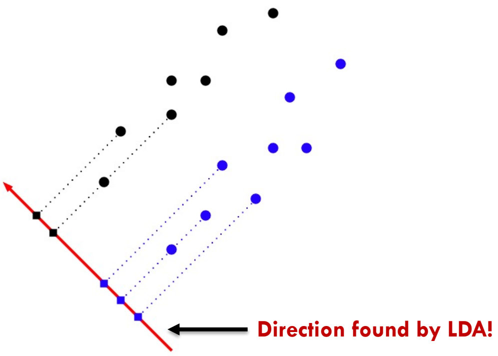
Linearly separable data

14

Linear Discriminant Analysis (LDA)

15

- Solution: A DR technique that tries to preserve the discriminatory information between different classes of the data set



15

Linear Discriminant Analysis

16

- Key Idea
  - Select the best direction for projection that
    - Maximizes their between-class separability while
    - Minimizing their within-class variability

- Key Requirements for Classification
  - Each class is as tight (compact) as possible
  - Their centroids are as far from each other as possible

16

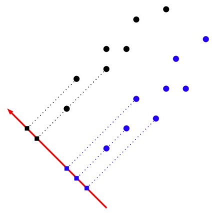
Linear Discriminant Analysis

17

Select the best direction for projection that

- Maximizes their between-class separability while
- Minimizing their within-class variability

- How do we quantify the separation between the two classes?
  - Measure the distance between the two class means in the projected space!



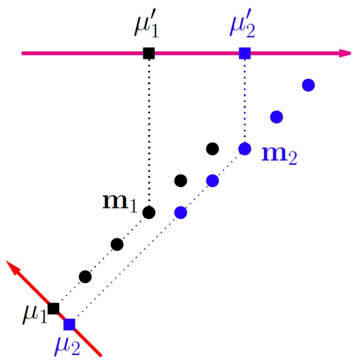
17

Linear Discriminant Analysis

18

How do we quantify the separation between the two classes?

- Measure the distance between the two class means in the projected space!
- This criterion alone might fail!



18

Linear Discriminant Analysis

19

How do we quantify the within-class variability?

- Compute variances of the projected classes

19

Linear Discriminant Analysis

20

- Linear Discriminant Analysis/ Fisher Discriminant Analysis
  - Objective: Perform dimensionality reduction while preserving as much of the class discriminatory information as possible.
    - Directions along which classes are best separated
- Find a projection $\mathbf{w}$ that
  - Maximizes the between-class separability
  - Minimize within-class variability
- Maximize Fisher's criteria

LDA

20

Linear Discriminant Analysis

21

- Linear Transformation (projection)

$$y = w^T x$$

- Fisher's criteria

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}}$$

Measure of between-class separability

Total within-class variance/scatter

$$|\tilde{m}_1 - \tilde{m}_2| = |w^T (m_1 - m_2)|$$

$$\tilde{s}_i^{\,2} = \sum_{y \in y_i} (y - \tilde{m}_i)^2$$

LDA

21

Linear Discriminant Analysis

22

- Fisher's criteria as explicit function of w

$$J(w) = \frac{w^T S_B w}{w^T S_w w}$$

$$\boxed{S_w^{-1} S_B w = \lambda w}$$

Eigenvalue problem!

The projection vector w is the eigenvector of $S_W^{-1} S_B$

LDA

22

Linear Discriminant Analysis

23

- The final step
  - Select the eigenvector that corresponds to the maximum eigenvalue to maximize class separability

LDA

23

Linear Discriminant Analysis

24

Algorithm

1. Mean normalization
2. Compute mean vectors $\boldsymbol{m}_i \in \mathbb{R}^D$ for all k classes
3. Compute scatter matrices S_W, S_B
4. Compute eigenvectors and eigenvalues of $S_W^{-1}S_B$
5. Select k eigenvectors $\boldsymbol{w}_i$ with the largest eigenvalues to form a $D \times k$-dimensional matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k]$
6. Project samples onto the new subspace using $\boldsymbol{W}$ and compute the new coordinates as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}$

- $\boldsymbol{X} \in \mathbb{R}^{n \times D}$: ith row represents the ith sample
- $\boldsymbol{Y} \in \mathbb{R}^{n \times k}$: Coordinate matrix of the n data points w.r.t. eigenbasis $\boldsymbol{W}$ spanning the k-dimensional subspace

LDA

24

LDA Example

25

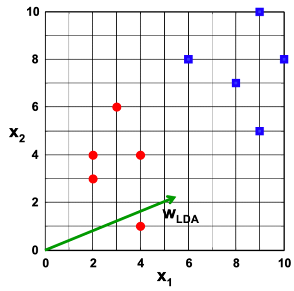
- **Compute the Linear Discriminant projection for the following two-dimensional dataset**
  - X1=(x_1,x_2)={(4,1),(2,4),(2,3),(3,6),(4,4)}
  - X2=(x_1,x_2)={(9,10),(6,8),(9,5),(8,7),(10,8)}
- **SOLUTION (by hand)**
  - The class statistics are:

$$S_1=\begin{bmatrix}0.80 & -0.40\\ -0.40 & 2.60\end{bmatrix};\ S_2=\begin{bmatrix}1.84 & -0.04\\ -0.04 & 2.64\end{bmatrix}$$

$$\mu_1=[3.00\quad 3.60];\qquad \mu_2=[8.40\quad 7.60]$$

  - The within- and between-class scatter are

$$S_B=\begin{bmatrix}29.16 & 21.60\\ 21.60 & 16.00\end{bmatrix};\ S_W=\begin{bmatrix}2.64 & -0.44\\ -0.44 & 5.28\end{bmatrix}$$



  - The LDA projection is then obtained as the solution of the generalized eigenvalue problem

$$\begin{bmatrix}11.89 & 8.81\\ 5.08 & 3.76\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix}=15.65\begin{bmatrix}v_1\\ v_2\end{bmatrix}\Rightarrow\begin{bmatrix}v_1\\ v_2\end{bmatrix}=\begin{bmatrix}0.91\\ 0.39\end{bmatrix}$$

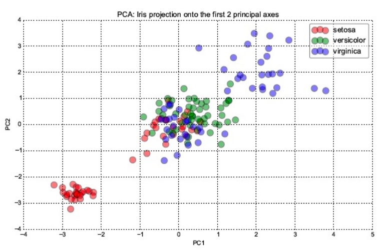
LDA

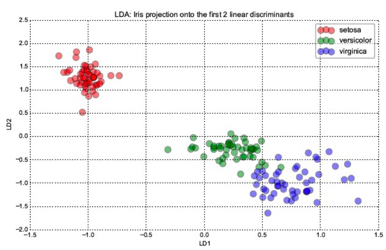
25

PCA versus LDA

26

- Like PCA, LDA looks at an eigenvalue problem for dimensionality reduction
- LDA: Eigenvalues indicate the importance of the corresponding direction with respect to classification performance



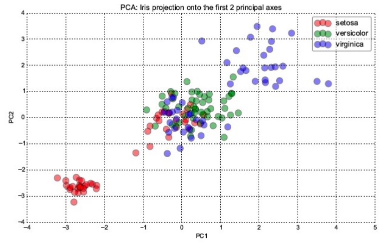


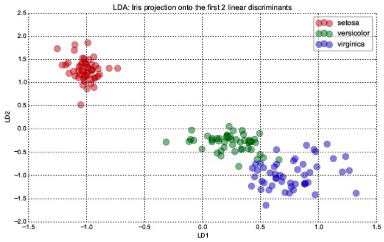
LDA

26

PCA versus LDA

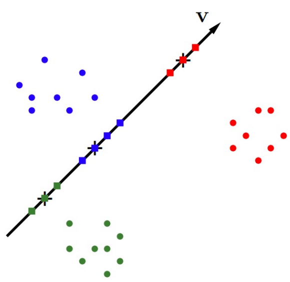
- PCA: Eigenvalues indicate the importance of the corresponding direction with respect to minimizing reconstruction error





Extension to Multiclass

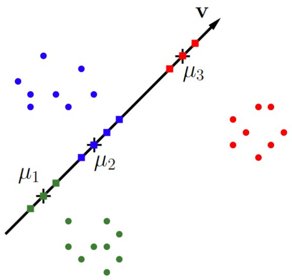
- Idea remains the same
  - Tightness of the projected classes is still described by the total within-class scatter



Extension to Multiclass

29

- Idea remains the same
- To maximize between-class separability - maximize the variance of the centroids



LDA

29

References

30

- R. O. Duda, P. E. Hart and D. G. Stork, **Pattern Classification.** 2nd edition, Wiley-Interscience publication.
- A. Martinez, A. Kak, "**PCA versus LDA**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

LDA

30

Thank You!

31