CS623: Introduction to Computing with Neural Nets *(lecture-16)*

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

Principal Component Analysis

Eaample: *IRIS Data (only 3 values out of 150)*

ID	Petal Length (a_1)	Petal Width (a_2)	Sepal Length (a_3)	Sepal Width (a_4)	Classification
001	5.1	3.5	1.4	0.2	Iris-setosa
051	7.0	3.2	4.7	1.4,	Iris-versicolor
101	6.3	3.3	6.0	2.5	Iris-virginica

Training and Testing Data

- Training: 80% of the data; 40 from each class: total 120
- Testing: Remaining 30
- Do we have to consider all the 4 attributes for classification?
- Do we have to have 4 neurons in the input layer?
- Less neurons in the input layer may reduce the overall size of the n/w and thereby reduce training time
- It will also likely increase the generalization performance (Occam Razor Hypothesis: *A simpler hypothesis (i.e., the neural net) generalizes better*

The multivariate data

$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	$\mathbf{X_5} \ldots \mathbf{X_p}$
x_{11}	x_{12}	x_{13}	x_{14}	$x_{15} \ldots x_{1p}$
x_{21}	x_{22}	x_{23}	x_{24}	$x_{25} \ldots x_{2p}$
x_{31}	x_{32}	x_{33}	x_{34}	$x_{35} \ldots x_{3p}$
x_{41}	x_{42}	x_{43}	x_{44}	$x_{45} \ldots x_{4p}$
			$\ldots$	
			$\ldots$	
x_{n1}	x_{n2}	x_{n3}	x_{n4}	$x_{n5} \ldots x_{np}$

Some preliminaries

- Sample mean vector: $<\mu_1, \mu_2, \mu_3, \ldots, \mu_p>$

  For the i^{th} variable: $\mu_i = (\Sigma^n_{j=1} x_{ij})/n$
- Variance for the i^{th} variable:

  $$\sigma_i^{\,2} = [\Sigma^n_{j=1} (x_{ij} - \mu_i)^2] / [n-1]$$
- Sample covariance:

  $$c_{ab} = [\Sigma^n_{j=1} ((x_{aj} - \mu_a)(x_{bj} - \mu_b))] / [n-1]$$

This measures the correlation in the data

In fact, the correlation coefficient

$$r_{ab} = c_{ab} / \sigma_a \sigma_b$$

Standardize the variables

- For each variable x_{ij}

  Replace the values by

$$y_{ij} = (x_{ij} - \mu_i)/\sigma_i^{\,2}$$

Correlation Matrix

$$R = \begin{bmatrix} 1 \;\; r_{12} \;\; r_{13} \ldots \; r_{1p} \\ r_{21} \;\; 1 \;\; r_{23} \cdots \; r_{2p} \\ \vdots \\ r_{p1} \;\; r_{p2} \;\; r_{p3} \cdots 1 \end{bmatrix}$$

Short digression: Eigenvalues and Eigenvectors

$AX=\lambda X$

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots a_{1p}x_p = \lambda x_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots a_{2p}x_p = \lambda x_2$$

$$\ldots$$

$$\ldots$$

$$a_{p1}x_1 + a_{p2}x_2 + a_{p3}x_3 + \ldots a_{pp}x_p = \lambda x_p$$

Here, λs are eigenvalues and the solution

$$<x_1, x_2, x_3, \ldots x_p>$$

For each λ is the eigenvector

Short digression: To find the Eigenvalues and Eigenvectors

Solve the characteristic function

$\det(A - \lambda I)=0$

Example:

$$\begin{pmatrix} -9 & 4 \\ 7 & -6 \end{pmatrix}$$

Verify:

$$\begin{pmatrix} -9 & 4 \\ 7 & -6 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = -13\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Characteristic equation

$(-9-\lambda)(-6-\lambda)-28=0$

Real eigenvalues: -13, -2

Eigenvector of eigenvalue -13:

(-1, 1)

Eigenvector of eigenvalue -2:

(4, 7)

$$I= \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

Next step in finding the PCs

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & \ldots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ & & \vdots & & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{bmatrix}$$

Find the eigenvalues and eigenvectors of *R*

Example

(from “Multivariate Statistical Methods: A Primer, by Brian Manly, 3rd edition, 1944)

49 birds: 21 survived in a storm and 28 died.
5 body characteristics given
X_1: body length; X_2: alar extent; X_3: beak and head length
X_4: humerus length; X_5: keel length
Could we have predicted the fate from the body charateristic

$$R = \begin{bmatrix} 1.000 & & & & \\ 0.735 & 1.000 & & & \\ 0.662 & 0.674 & 1.000 & & \\ 0.645 & 0.769 & 0.763 & 1.000 & \\ 0.605 & 0.529 & 0.526 & 0.607 & 1.000 \end{bmatrix}$$

Eigenvalues and Eigenvectors of *R*

Component	Eigen value	First Eigen-vector: V_1	V_2	V_3	V_4	V_5
1	3.612	0.452	0.462	0.451	0.471	0.398
2	0.532	-0.051	0.300	0.325	0.185	-0.877
3	0.386	0.691	0.341	-0.455	-0.411	-0.179
4	0.302	-0.420	0.548	-0.606	0.388	0.069
5	0.165	0.374	-0.530	-0.343	0.652	-0.192

Which principal components are important?

- Total variance in the data=

  $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5$

  = sum of diagonals of *R*= *5*
- First eigenvalue= 3.616 ≈ 72% of total variance 5
- Second ≈ 10.6%, Third ≈ 7.7%, Fourth ≈ 6.0% and Fifth ≈ 3.3%
- ***First PC is the most important and sufficient for studying the classification***

Forming the PCs

- $Z_1 = 0.451X_1 + 0.462X_2 + 0.451X_3 + 0.471X_4 + 0.398X_5$
- $Z_2 = -0.051X_1 + 0.300X_2 + 0.325X_3 + 0.185X_4 - 0.877X_5$
- For all the 49 birds find the first two principal components
- This becomes the new data
- Classify using them

For the first bird

X_1=156, X_2=245, X_3=31.6, X_4=18.5, X_5=20.5

After standardizing

Y_1=(156-157.98)/3.65=-0.54,

Y_2=(245-241.33)/5.1=0.73,

Y_3=(31.6-31.5)/0.8=0.17,

Y_4=(18.5-18.46)/0.56=0.05,

Y_5=(20.5-20.8)/0.99=-0.33

PC_1 for the first bird=

Z_1= 0.45X(-0.54)+ 0.46X(0.725)+0.45X(0.17)+0.47X(0.05)+0.39X(-0.33)

=0.064

Similarly, Z_2= 0.602

Reduced Classification Data

- Instead of

X_1	X_2	X_3	X_4	X_5
←		49 rows		→

- Use

Z_1	Z_2
← 49	rows →

Other Multivariate Data Analysis Procedures

- Factor Analysis
- Discriminant Analysis
- Cluster Analysis

To be done gradually