CS423: Data mining

Principle Component Analysis Lab

Jakramate Bootkrajang

Department of Computer Science

Chiang Mai University

Principle Component Analysis

- Standardising X.
  - Subtract feature value by its mean and divide by its S.D.
- Calculate V, a covariance matrix for X
  - Useful function `cov(X)`
- Find all the eigenvectors of V.
  - Useful function `eigvecs(V)` and `eigvals(V)`.
- Select k most important principle components put it in a matrix A.
- Project X onto A by calculating AX^T.

The `eigvecs(V)` and `eigvals(V)` function.

- For computing eigenvectors and eigenvalues of an input matrix. They are part of **`LinearAlgebra`** package.
- **`eigvecs(V)`** returns return a matrix A whose columns are the eigenvectors of V
- `A = eigvecs([1.0 0.0 0.0; 0.0 3.0 0.0; 0.0 0.0 18.0])`
- λ `= eigvals([1.0 0.0 0.0; 0.0 3.0 0.0; 0.0 0.0 18.0])`
- Type **`?eigvals`** or **`?eigvecs`** to see the help pages

Reading CSV file

- Install package
  - `using Pkg; Pkg.add("CSV")`
- Use package
  - `using CSV`
- Reading CSV data file
  - `m = CSV.read("filename.csv")`

Visualising m-dimensional data using PCA

```
using CSV
d = CSV.read("ionosphere.csv")  # reading data
# d is DataFrame which can be indexed by column


# z-score normalisation
X = convert(Array,[d[1] d[3:34]])
X = (X - repeat(mean(X,dims=1),351,1))
        ./  repeat(std(X,dims=1), 351, 1)
```

Visualising m-dimensional data using PCA [2]

```
# continuing

# calculating covariance

V = cov(X)

# finding eigenvectors and eigen values

A = eigvecs(V)

\lambda = eigvals(V)
```

Visualising m-dimensional data using PCA [3]

```
# eigenvectors and eigenvalues are sorted
# in ascending order, first PC is the last one
# projecting the data onto 2D space
# defined by the first two principle components
P = X * A[:, [32,33]]
```

Visualising m-dimensional data using PCA [4]

```
# find index of class -1 and 1
Y = d[35] .== "g"
# creating scatter plot
using Plots
scatter(P[Y.==false,1], P[Y.==false,2])
scatter!(P[Y.==true,1], P[Y.==true, 2])
```

Reference

- Learn X in Y minutes for Julia

`https://learnxinyminutes.com/docs/julia/`