# Tutorial 4

# Principal Component Analysis for Feature Dimensionality Reduction

# BT3017

Due date: 21st Feb 2021 (Monday) 2359 hrs

Semester 2, AY21/22, School of Computing, National University of Singapore

***IMPORTANT:***

*For this tutorial, you are supposed to submit your project file to LUMINUS.*

*Instruction for submission:*

- *Create a folder using the following naming convention:*

  *StudentNumber_yourName_Tut3*

- *Put your python file and also the results in this folder.*

- *Zip your folder. Name your zip file using the following convention:*

  *StudentNumber_yourName_Tut3.zip*

  *For example, if your student number is A1234567B, and your name is Chow Yuen Fatt, for this tutorial, your file name should be A1234567B_ChowYuenFatt_Tut3.zip*

- *Submit the zip file in the "Tutorial-3 Submit Here" folder in Luminus.*

The objectives of this tutorial is to understand Principal Component Analysis and its use in feature dimensionality reduction.

Q1

Write a python program to do the following:

- Read the file “seeds_with_headers.csv” into a pandas dataframe. Note that “seeds_with_headers.csv” was modified from “wheat-seeds.csv” downloaded from

  https://machinelearningmastery.com/implement-backpropagation-algorithm-scratch-python/

  Note that there are 8 features arranged in 8 columns i.e. feat1 to feat8.

  Note also that each row is one data point.

- import numpy as np
- Use np.mean to compute the mean of each of the 8 features
- Use np.cov to compute the covariance matrix of the data in the file. *hint: set rowvar = False*
- Use np.linalg.eig to do eigen decomposition of the covariance matrix.
- Use np.argsort to sort the eigenvalues in descending order.
- Use np.dot to show that the eigenvectors are orthogonal to each other.

Q2

Project each data point onto all the eigenvectors and store the results of projection into a matrix coef of size number_of_data_points by 8.

- Reconstruct the data points using coef and the eigenvectors.
- Calculate the squared reconstruction errors.
- Print the squared reconstruction error for each data point.

Q3

Repeat Q2 by projecting each data point onto the 4 eigenvectors corresponding to the 4 biggest eigenvalues.

Q4

Repeat Q2 by projecting each data point onto the 4 eigenvectors corresponding to the 4 smallest eigenvalues.

Q5

Compare and comment on the squared reconstruction errors in Q2, Q3, and Q4.