วิชา PCA + KNN
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

Given a glass identification dataset with 214 instances and 10 attributes. Here is information of each attribute:

- Id number: 1 to 214
- RI: refractive index
- Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
- Mg: Magnesium
- Al: Aluminum
- Si: Silicon
- K: Potassium
- Ca: Calcium
- Ba: Barium
- Fe: Iron
- Type of glass: (class attribute)
  - -- 1 building_windows_float_processed
  - -- 2 building_windows_non_float_processed
  - -- 3 vehicle_windows_float_processed
  - -- 4 vehicle_windows_non_float_processed (none in this database)
  - -- 5 containers
  - -- 6 tableware
  - -- 7 headlamps

1) Construct the classifier to predict type of glass with-out the use of PCA

Pre-processing:

- Show each step of your pre-processing

Classification:

- Create a classification model with K-NN operator to predict Type of glass.
- Use Split-Data operator (0.7 for training and 0.3 for testing) with local random seed = 1992
- Use Euclidean distance on numerical variables
- What is best K to obtain best accuracy => Show results on different values of K
- Show confusion-matrix for each class

2) Construct the classifier to predict type of glass with the use of PCA

Pre-processing:

- Show each step of your pre-processing (include PCA)
- Show & explain setting of PCA parameters
- Explain output of PCA in terms of
    - Principal components
    - Accumulate variance

Classification:

- Show format of the dataset in terms of new dimensions
- Create a classification model with K-NN operator to predict Type of glass.
- Use Split-Data operator (0.7 for training and 0.3 for testing) with local random seed = 1992
- Use Euclidean distance on numerical variables
- What is best combination of PCA (variance) & K to obtain best accuracy
- Show confusion matrix for each class

3) Discuss and Compare accuracy obtained between with PCA and without PCA
4) Submit => student_ID_glass_KNN.xml
   => student_ID_glass_PCA_KNN.xml

5) Extra-points: Redo all tasks using Python or R
   Submit => student_ID_glass_KNN.ipynb or py or R