



## Exercise: PCA+ KNN

### ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

#### Exercise : การจำแนกประเภทของไวน์ (wine classification)

ตารางต่อไปนี้แสดงเอาต์พุตของการวิเคราะห์องค์ประกอบหลัก (PCA) ของข้อมูลไวน์ (non-normalized) ซึ่งตัวแปรแสดงถึงลักษณะทางเคมีของไวน์ และแต่ละกรณีคือไวน์ที่แตกต่างกัน

	Principal Components					Std Dev
	1	2	3	4	5	
Alcohol	0.001	0.013	0.014	-0.030	0.129	0.8
MalicAcid	-0.001	0.009	0.167	-0.427	-0.402	1.2
Ash	0.000	-0.002	0.054	-0.009	0.006	0.3
Ash_Alcalinity	-0.004	-0.045	0.976	0.176	0.060	3.6
Magnesium	0.014	-0.998	-0.040	-0.031	0.006	14.7
Total Phenols	0.001	0.002	-0.015	0.164	0.316	0.7
Flavanoids	0.002	0.000	-0.049	0.214	0.545	1.1
Nonflavanoid_Phenols	0.000	0.002	0.004	-0.025	-0.040	0.1
Proanthocyanins	0.001	-0.007	-0.031	0.082	0.244	0.7
Color Intensity	0.002	0.022	0.097	-0.804	0.536	1.6
Hue	0.000	-0.002	-0.021	0.096	0.064	0.2
OD280/OD315	0.001	-0.002	-0.022	0.220	0.261	0.7
Proline	1.000	0.014	0.004	0.001	-0.004	351.5
Variance	123594.453	194.345	11.424	2.388	1.391	
% Variance	99.830%	0.157%	0.009%	0.002%	0.001%	
Cumulative %	99.830%	99.987%	99.996%	99.998%	99.999%	

- พิจารณาแถวใกล้ด้านล่างที่ระบุว่า "ความแปรปรวน" อธิบายว่าเหตุใดความแปรปรวนของคอลัมน์ 1 จึงมากกว่าคอลัมน์อื่นมาก?
- จงให้เหตุผลเกี่ยวกับการใช้การทำให้เป็นมาตรฐาน (normalization)?
- เปรียบเทียบและแสดงความคิดเห็นผลลัพธ์ระหว่าง (Normalization + PCA) และ (Non-Normalization + PCA) มีผลลัพธ์แตกต่างกันอย่างไร?

d) สร้างโมเดลทำนายประเภทของไวน์ โดยใช้เทคนิค K-NN

**Pre-processing:**

- Show each step of your pre-processing including step of your feature selection

**Classification:**

- Create a classification model with K-NN operator to predict Type of wine.
- Use Split-Data operator (0.7 for training and 0.3 for testing)
- Show results(accuracy) on different values of K, then identify best K

e) Repeat the previous steps with the use of PCA

**PCA:**

- Show your steps of 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 (components)
- Create a classification model with K-NN operator to predict Type of wine.
- Use Split-Data operator (0.7 for training and 0.3 for testing)
- Show results(accuracy) on different values of K, then identify best K

f) Discuss and Compare accuracy obtained between with PCA and without PCA

Submit      => student\_ID\_wine\_KNN.xml  
                  => student\_ID\_wine\_PCA\_KNN.xml  
                  => student\_ID\_answer.pdf

g) Redo all tasks using Python

Submit   => student\_ID\_wine.ipynb