

## **LECTURE 8-3**

# **MACHINE LEARNING**

**DR. PRAPASSORN TANTIPHANWADI**

**INDUSTRIAL ENGINEERING, FACULTY OF ENGINEERING AT KHAMPAENGSSEN**

**DECEMBER 2565**

# CONTENT

- What is Machine Learning?
- Type of Machine Learning?
- Overfitting and Underfitting
- ML Algorithms

- 1) Simple Linear Regression
- 2) Multiple Linear Regression
- 3) Logistic Regression
- 4) Decision Tree
- 5) Random Forest

- 6) Support Vector Machine (SVM)
- 7) Naïve Bayes
- 8) K-NN
- 9) PCA
- 10) K-Mean Clustering

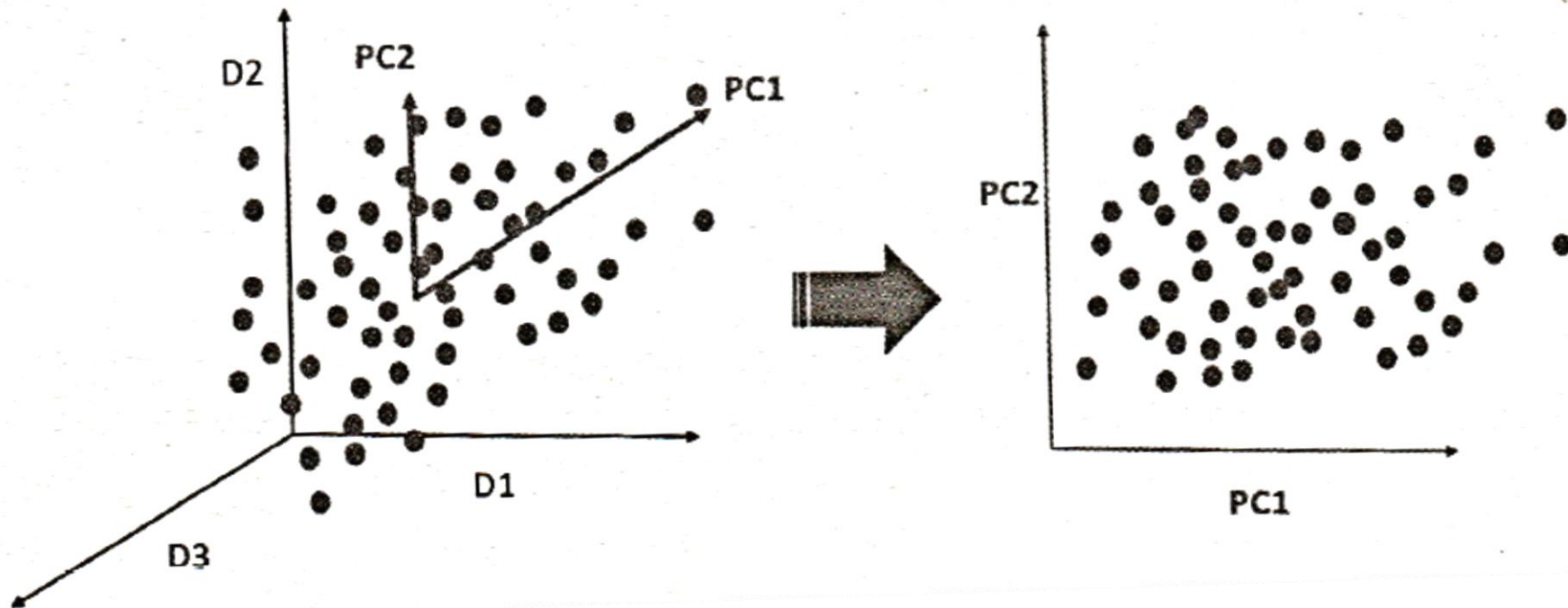
## การลดมิติข้อมูลด้วย PCA

PCA ย่อมาจาก **Principle Component Analysis** มีชื่อในภาษาไทยว่า “การวิเคราะห์องค์ประกอบหลัก” เป็นกระบวนการทำงานโดยใช้หลักการทางสถิติและคณิตศาสตร์เพื่อสร้างตัวแปรในการเป็น “ตัวแทน” ของชุดคอลัมน์ของข้อมูล ซึ่งตัวแปรที่เป็นตัวแทนนี้จะเสมือนเป็นการย่อข้อมูลมุมมองหนึ่งของความสัมพันธ์ของหลายๆ คอลัมน์ในชุดข้อมูลที่มีความสัมพันธ์กัน

PCA มีประโยชน์ในการลดจำนวนมิติข้อมูลหรือคอลัมน์เพื่อนำไปใช้ในการสร้างโมเดลเพื่อวิเคราะห์ ซึ่งการลดอาจทำให้ตัดคอลัมน์ข้อมูลที่ไม่จำเป็นต่อการวิเคราะห์หรือคอลัมน์ที่ซ้ำซ้อนออกไป ทำให้ลดเวลาในการสร้างโมเดลและการทำนาย ทำให้ PCA มักใช้กันบ่อยกับชุดข้อมูลที่มีคอลัมน์หรือจำนวน feature เป็นจำนวนมาก เช่น รูปภาพ ซึ่งใช้ในการทำโมเดลสำหรับงาน Face Recognition หรือ Image Recognition

# PCA

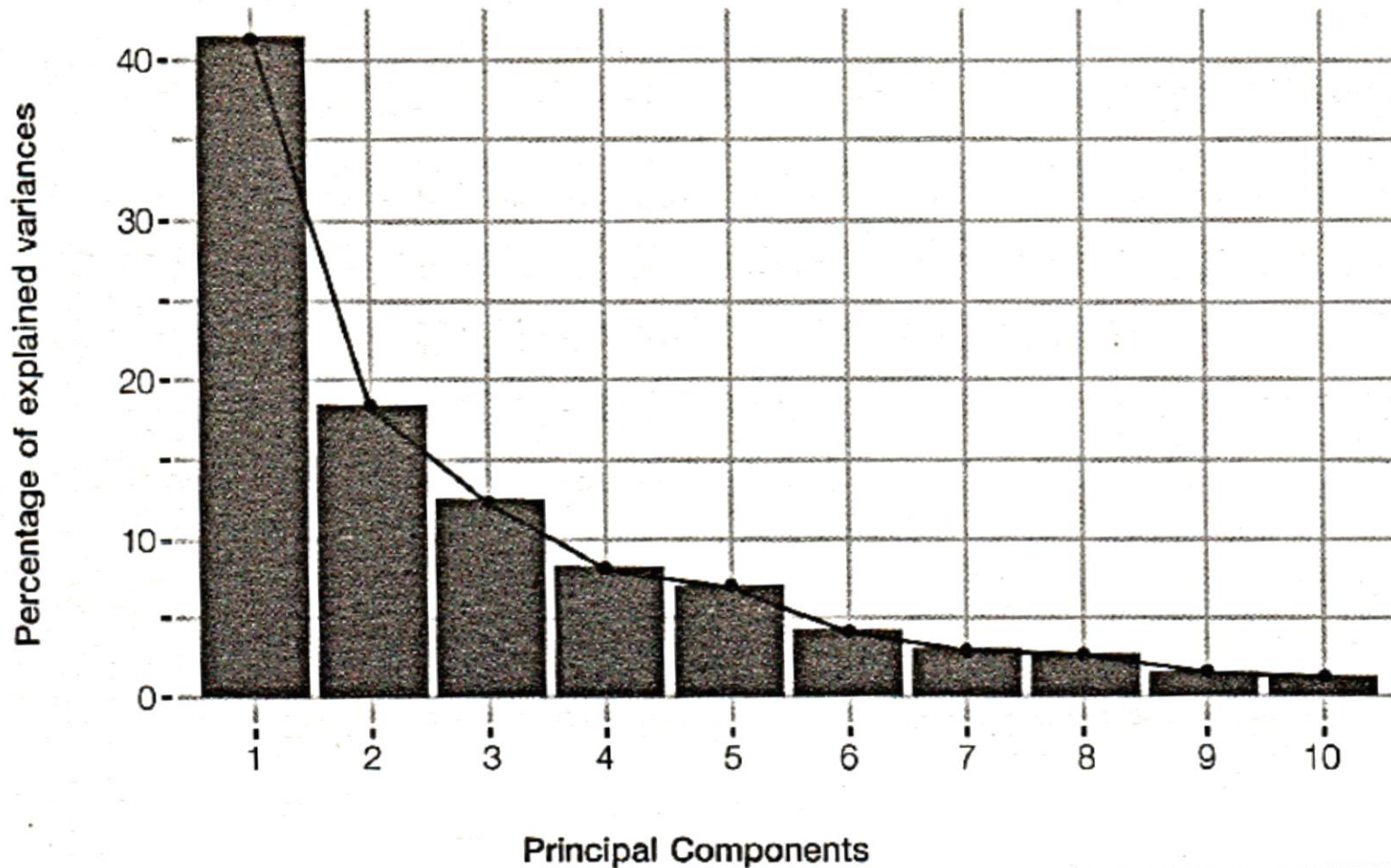
ขั้นตอนการทำงานของ PCA คือ การตีเส้นเพื่อสร้างระนาบใหม่จากจุดกึ่งกลางของข้อมูล เพื่อใช้เป็นตัวแทนในการอธิบายความสัมพันธ์กับแต่ละ Data point โดยเส้นนี้จะคำนวณโดยใช้หลักการคำนวณของ Linear Regression ด้วย least squares ผลลัพธ์ที่ได้ คือ จะเกิดเส้นหรือที่เรียกว่า Principal Component ขึ้นมาหลาย ๆ ตัวเพื่อแทนชุดข้อมูล ทำให้ข้อมูลถูกปรับมุมมองข้อมูลมาเป็นข้อมูลตามแกนของ Principal component หากข้อมูลเดิมมีอยู่  $d$  มิติ จะถูกลดเหลือ  $k$  มิติ โดย  $d > k$





# PCA

จากรูป เดิมชุดข้อมูลมีอยู่ 3 มิติ จะลดเหลือเพียง 2 มิติตามแกนของ Principal component และแต่ละ component จะไม่มีความสัมพันธ์ ซึ่ง component ตัวแรกที่เราสร้างจะมีความหลากหลายของข้อมูล (variance) มากที่สุด จึงถือว่าเป็นตัวแทนชุดข้อมูลจำนวนมากที่สุด เนื่องจากมีการกระจายตัวของข้อมูลมากที่สุด ส่วน component ถัดมาจะมีค่าความหลากหลายของข้อมูลลดลงเรื่อยๆ ดังรูป

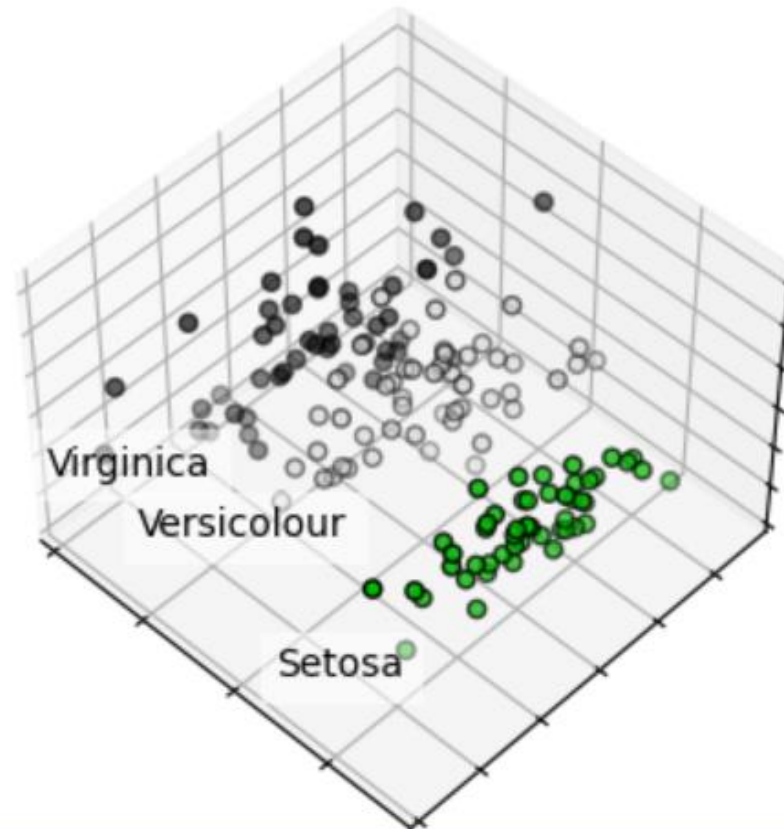


# PCA

## PCA example with Iris Data-set

Principal Component Analysis applied to the Iris dataset.

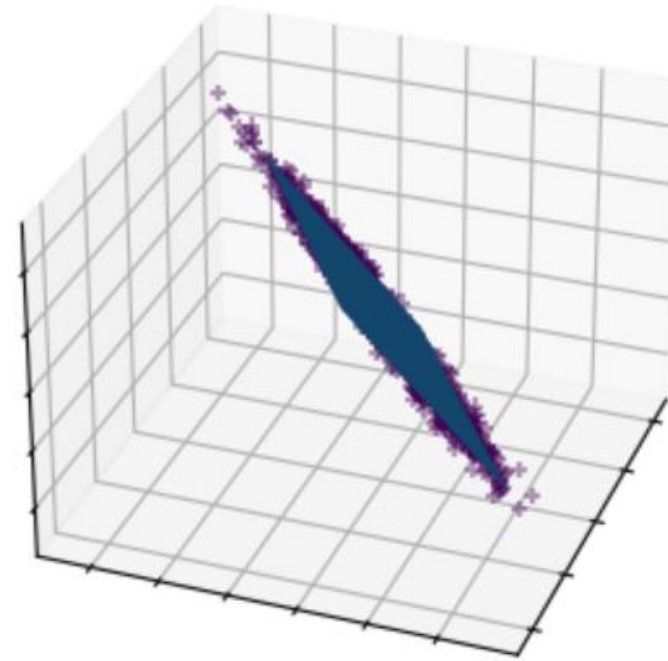
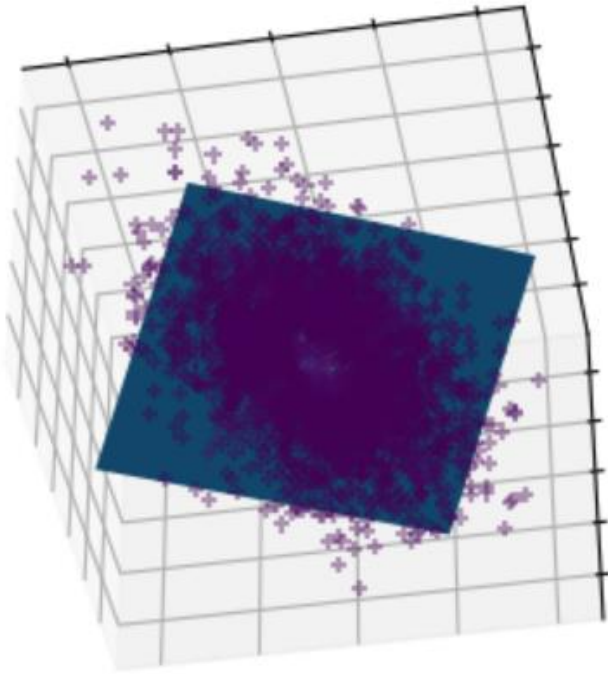
See [here](#) for more information on this dataset.



# PCA

[https://scikit-learn.org/stable/auto\\_examples/decomposition/plot\\_pca\\_3d.html#sphx-glr-auto-examples-decomposition-plot-pca-3d-py](https://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_3d.html#sphx-glr-auto-examples-decomposition-plot-pca-3d-py)

These figures aid in illustrating how a point cloud can be very flat in one direction—which is where PCA comes in to choose a direction that is not flat.





# K-NEAREST NEIGHBOR (K-NN)

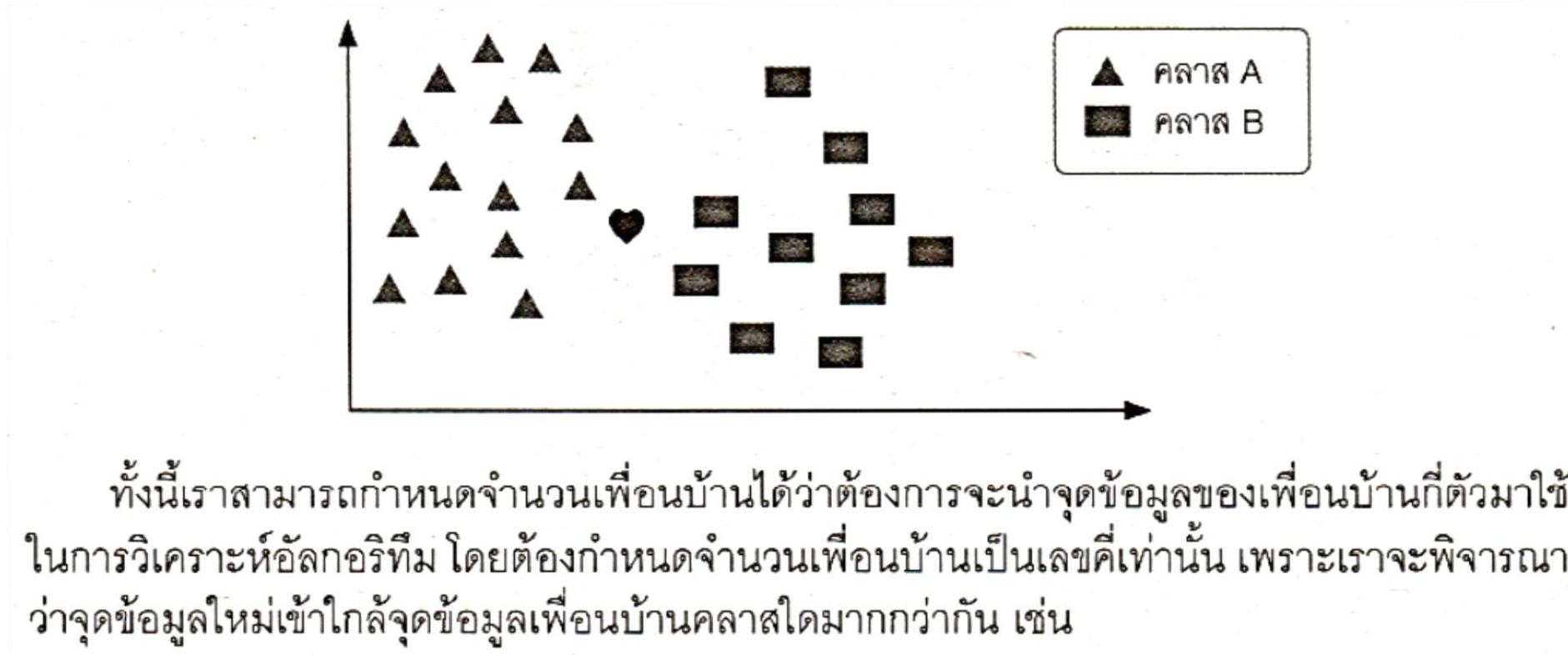
K-NN เป็นอัลกอริทึมหนึ่งของ Machine Learning ที่จัดอยู่ในประเภท Supervised Learning อัลกอริทึมนี้จะใช้ข้อมูลของ Dataset ทั้งหมดโดยไม่แตกข้อมูลออกเป็น Training data และ Testing data

เริ่มแรกก็นำข้อมูลมาพล็อตกราฟแผนภูมิการกระจาย (Scatter Plots) ก่อน ในที่นี้ข้อมูลแบ่งเป็น 2 คลาส คือ คลาส A แทนด้วยสามเหลี่ยม และคลาส B แทนด้วยสี่เหลี่ยม เมื่อมีข้อมูลเข้ามาใหม่ (ในที่นี้แทนด้วยรูปหัวใจ) และต้องการทำนายว่าข้อมูลใหม่นั้นอยู่ในคลาส A หรือ คลาส B เราสามารถทำนายได้โดยใช้อัลกอริทึม K-NN

ชื่อก็บอกอยู่แล้วว่า K-Nearest Neighbors หลักการทำงาน ก็คือ ที่มันจะทำนายว่าจุดข้อมูลใหม่ เข้าใกล้จุดข้อมูลของเพื่อนบ้านคลาส A หรือ คลาส B มากกว่ากัน ถ้าใกล้คลาส A ก็แปลว่า ข้อมูลใหม่นั้นเป็นคลาส A แต่ถ้าใกล้คลาส B ก็แปลว่า ข้อมูลใหม่นั้นเป็นคลาส B

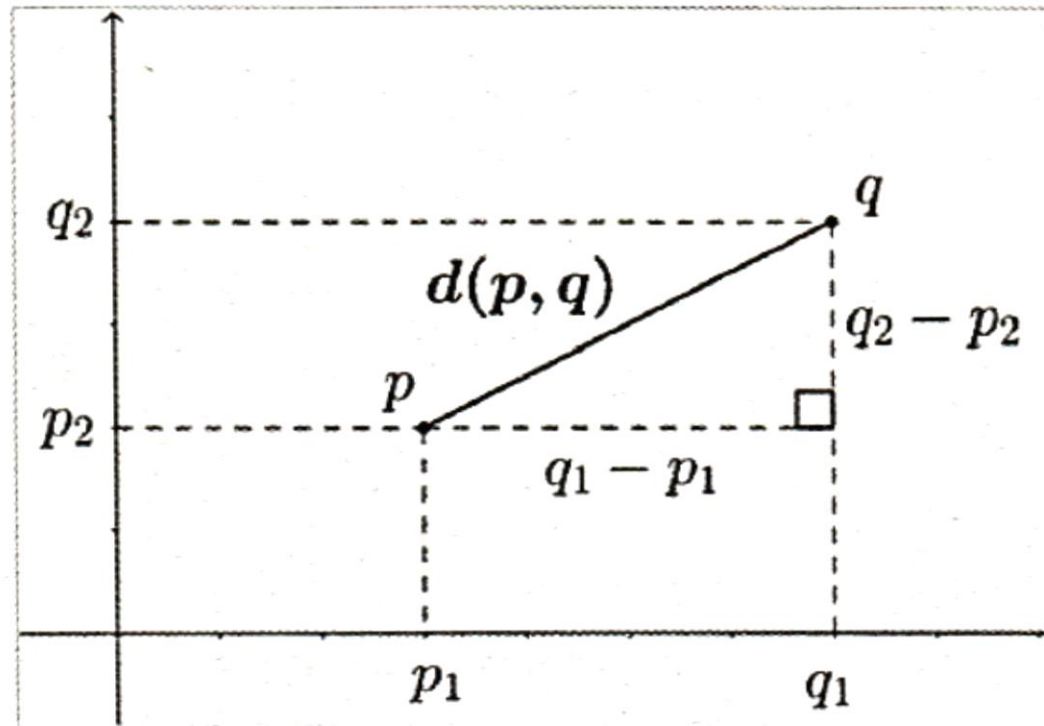


# K-NEAREST NEIGHBOR (K-NN)



# K-NEAREST NEIGHBOR (K-NN)

การหาระยะทางระหว่างแต่ละจุดบนแกน x-y แบบ 2 มิติว่ามีความห่างไกลหรือใกล้กันเท่าใด สามารถคำนวณได้ด้วยสูตร Euclidean distance ดังนี้



# K-NEAREST NEIGHBOR (K-NN)

$$d(p,q)^2 = \sqrt{(p - q)^2}$$

ส่วนการหาระยะทางระหว่างจุด 2 จุดบนแกนแบบหลายมิติ สามารถคำนวณได้ด้วยสูตรต่อไปนี้

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

ตัวอย่างของการนำอัลกอริทึม K-NN ไปใช้งาน เช่น

- Face recognition คือ ระบบการจดจำใบหน้า
- Recommender system คือ ระบบแนะนำสินค้าและบริการต่าง ๆ ตัวอย่างเช่น YouTube, Netflix นำข้อมูลลูกค้าใหม่มาเทียบกับข้อมูลลูกค้าเดิม เพื่อพิจารณาว่าลูกค้ามีพฤติกรรมชอบดูวิดีโอประเภทใด และทำการจัดกลุ่มลูกค้าที่มีพฤติกรรมคล้ายกันไว้ในกลุ่มเดียวกัน จากนั้นก็จะส่งวิดีโอแนะนำไปยังกลุ่มลูกค้าต่าง ๆ ได้ถูกต้องตรงตามความสนใจของลูกค้า

# K-NEAREST NEIGHBOR (K-NN)

## ตัวอย่าง : Image Recognition วิเคราะห์และทำนายรูปภาพด้วย K-NN

ตัวอย่างนี้จะเป็นการวิเคราะห์รูปภาพบุคคลสำคัญของโลก โดยอ่านไฟล์รูปภาพเพื่อมาเปรียบเทียบกับฐานข้อมูลรูปภาพ Labeled Faces in the Wild (LFW) people dataset ซึ่งเป็น Dataset ที่ไลบรารี scikit-learn ได้จัดเตรียมไว้แล้ว โดยจะมีรูปภาพขาวดำของบุคคลสำคัญต่างๆ อยู่ใน Dataset นี้ เช่น George W.Bush, Colin Powell เป็นต้น ซึ่งแต่ละบุคคลจะประกอบไปด้วยหลายภาพเพื่อนำมาใช้ในการเปรียบเทียบ

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch\\_lfw\\_people.html?highlight=lfw\\_people#](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_lfw_people.html?highlight=lfw_people#)

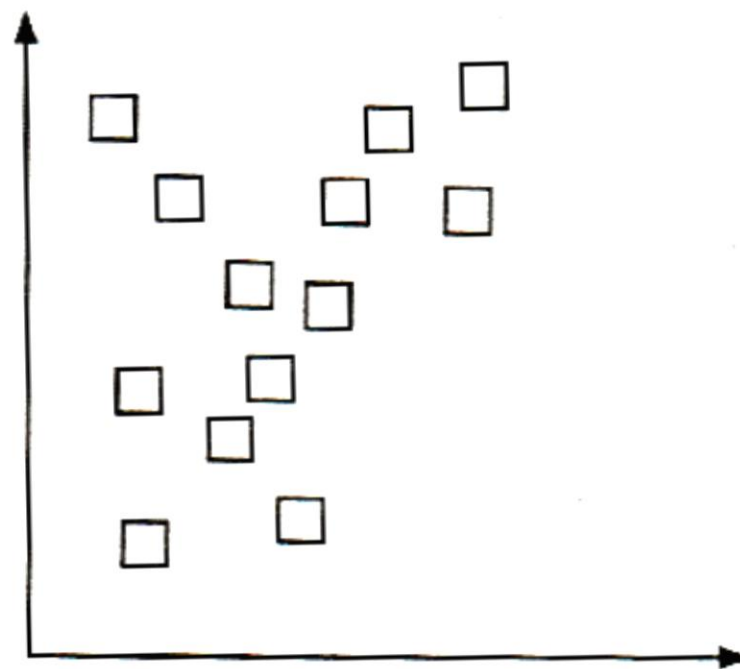


# K-MEAN

K-Means Clustering เป็นอัลกอริทึมหนึ่งของ Machine Learning ที่จัดอยู่ในประเภท Unsupervised Learning การทำนายผลของ K-Means จะไม่ได้ผลลัพธ์ออกมาเป็นเลเบล เพราะมันเรียนรู้แบบไม่มีผู้สอน

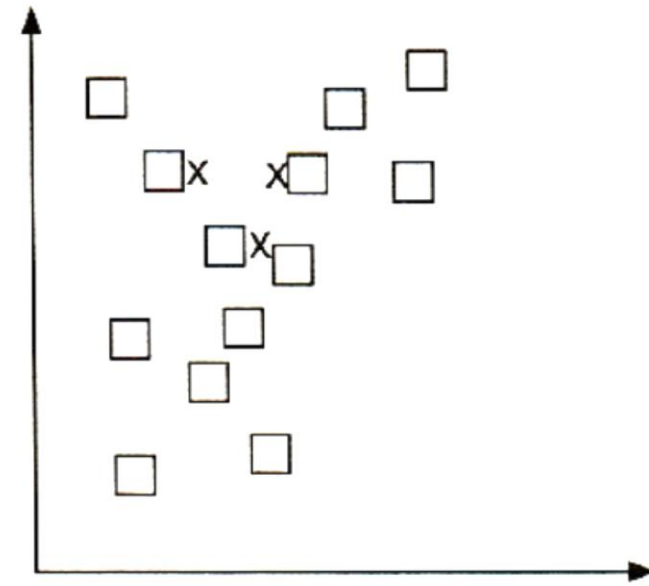
หลักการทำงานของอัลกอริทึมนี้ คือ จะทำการ Clustering จัดกลุ่มข้อมูลที่เหมือนกันหรือคล้ายกันไว้ในกลุ่มเดียวกัน ซึ่งการจัดกลุ่มมีขั้นตอนดังนี้

ขั้นตอนที่ 1 เอาจุดข้อมูลทั้งหมดมาพล็อตกราฟ แผนภูมิการกระจาย (Scatter Plots) ก่อน

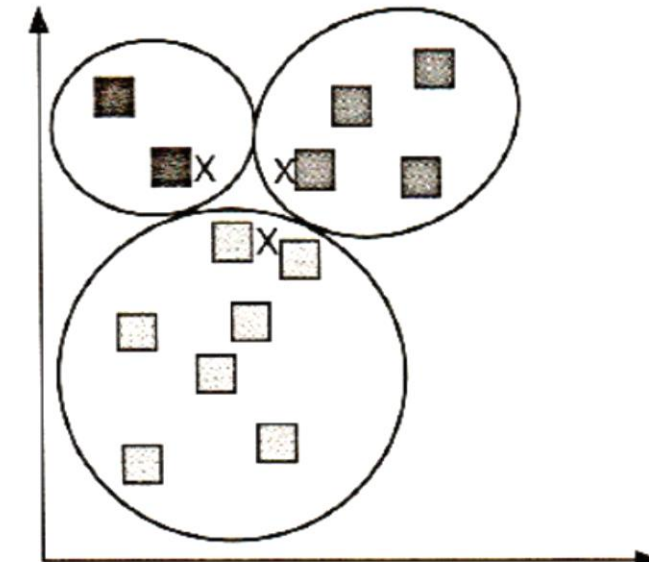


# K-MEAN

ขั้นตอนที่ 2 พิจารณาว่าเราต้องการแบ่งข้อมูลเป็นกี่กลุ่ม เช่นในที่นี้เราต้องการแบ่ง 3 กลุ่ม ก็กำหนด  $k=3$  จากนั้นสุ่มเลือกค่า Centroid ซึ่งเป็นค่าเฉลี่ย (ค่า Means) ของแต่ละกลุ่มออกมา ในที่นี้กำหนดว่าจะแบ่ง 3 กลุ่ม ดังนั้น สุ่มเลือกค่า Centroid มา 3 จุด

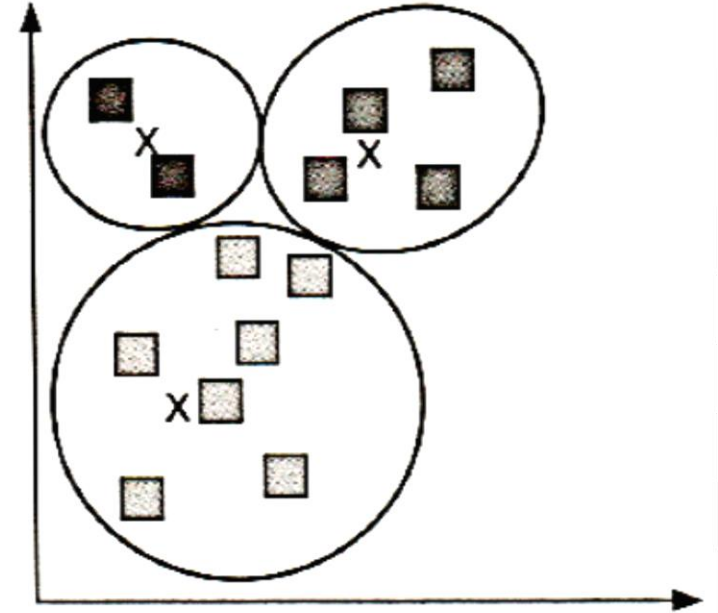


ขั้นตอนที่ 3 คำนวณหาระยะทางระหว่างจุดข้อมูล Data point แต่ละจุดกับ Centroid ด้วยสูตร Euclidean distance ถ้าจุดข้อมูลใดอยู่ใกล้ Centroid ใดก็จะถูกจัดอยู่ในกลุ่มเดียวกับ Centroid นั้น

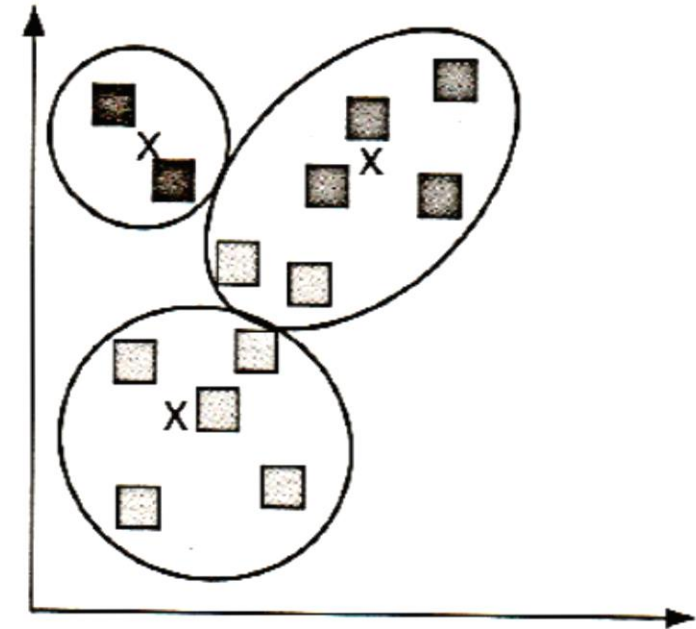


# K-MEAN

ขั้นตอนที่ 4 เมื่อจัดกลุ่มเสร็จแล้วก็จะทำการ  
คำนวณค่า Centroid ของแต่ละกลุ่มใหม่ โดยคำนวณ  
จากค่าเฉลี่ยของสมาชิกทุกตัวในกลุ่มนั้น



ขั้นตอนที่ 5 เมื่อได้ค่า Centroid ใหม่แล้ว ก็  
คำนวณหาระยะห่างระหว่างจุดข้อมูล Data point  
แต่ละจุดกับ Centroid เหมือนเดิม โกล้ Centroid  
ไหนก็จัดเข้ากลุ่มเดียวกับ Centroid นั้น



# K-MEAN

ตัวอย่างของการนำอัลกอริทึม K-Means Clustering ไปใช้งาน เช่น Customer Segmentation คือ การจับกลุ่มลูกค้าที่มีพฤติกรรมคล้ายกันไว้ในกลุ่มเดียวกันโดยอาจแบ่งการพิจารณาออกได้เป็นหลายประเภท

- Demographic information แบ่งกลุ่มลูกค้าจากเพศ อายุ สถานภาพ (โสด, สมรส, หย่า) รายได้ การศึกษา อาชีพ เป็นต้น
- Geographical information การแบ่งกลุ่มลูกค้าจาก Geographical ไม่มีกฎตายตัวแน่นอน ขึ้นกับว่าเราต้องการแบ่งกลุ่มลูกค้าแบบใด เช่น แบ่งตามจังหวัด แบ่งตามประเทศ เป็นต้น
- Psychographics แบ่งกลุ่มลูกค้าตามอุปนิสัยส่วนตัว ตาม lifestyle การใช้ชีวิต หรือ แบ่งตามสถานภาพทางสังคม เป็นต้น
- Behavioral data แบ่งกลุ่มลูกค้าตามพฤติกรรมการจับจ่ายใช้สอยโดยดูจากจำนวนรายการที่ซื้อ ราคาที่ซื้อ ช่องทางการซื้อ สินค้าที่ซื้อ เป็นต้น



# K-MEAN

## ตัวอย่าง : การทำ Customer Segmentation ด้วย K-Means Clustering

---

เราจะลองมาดูตัวอย่างการทำ Customer Segmentation เพื่อแบ่งกลุ่มลูกค้าของห้างสรรพสินค้าแห่งหนึ่งกัน

สมมติข้อมูลลูกค้าของห้างสรรพสินค้าแห่งหนึ่งเป็นดังนี้

**A demo of K-Means clustering on the handwritten digits data**

[https://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_kmeans\\_digits.html](https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_digits.html)

➤ IRIS dataset -- example

<https://medium.com/mlearning-ai/k-means-clustering-with-scikit-learn-e2af706450e4>

# HOMework

## 1. PCA classification:

<https://machinelearningknowledge.ai/complete-tutorial-for-pca-in-python-sklearn-with-example/>

id	gender	PPE	DFA	RFDE	numPulses	numPeric	meanPeric	stdDevPer	locPcctSite	locAbsSite	rapLitter	ppgLitter	ddpLitter	locShimmy	locDddShm	appShimmy	appShimw	appShimw	appShimw	ddaShimw
0	1	0.85247	0.71826	0.57227	240	239	0.008064	8.88E-05	0.00218	1.74E-05	0.00087	0.00129	0.002	0.05083	0.517	0.00011	0.01496	0.04838	0.09034	0.00034
0	1	0.76686	0.69461	0.53966	214	231	0.008248	7.31E-05	0.00191	1.61E-05	0.00052	0.00112	0.00017	0.05016	0.502	0.0032	0.02675	0.06195	0.06961	0.00061
0	1	0.85083	0.67604	0.58982	232	231	0.00834	6.94E-05	0.00176	1.47E-05	0.00057	0.00111	0.00171	0.09902	0.897	0.05094	0.06497	0.07772	0.15282	0.00028
1	0	0.81121	0.79672	0.59257	178	177	0.010858	0.000383	0.00493	4.53E-05	0.00149	0.00388	0.00446	0.05451	0.527	0.02395	0.02857	0.04662	0.07185	0.00028
1	0	0.8297	0.79792	0.53058	236	235	0.008162	0.000669	0.00351	4.71E-05	0.00146	0.00227	0.00499	0.0561	0.497	0.02909	0.03227	0.05278	0.08728	0.00028
1	0	0.5078	0.78744	0.65451	226	221	0.007931	0.000696	0.00783	5.97E-05	0.00232	0.00312	0.00697	0.07752	0.678	0.02805	0.04767	0.06451	0.11415	0.00028
2	1	0.7095	0.62145	0.54543	322	321	0.007991	0.001007	0.00221	1.13E-05	0.00056	0.00094	0.00139	0.02302	0.28	0.0155	0.01971	0.02274	0.0445	0.00028
2	1	0.83671	0.62079	0.51179	318	317	0.006074	0.000136	0.00282	1.71E-05	0.00034	0.00088	0.00103	0.063	0.539	0.02949	0.04091	0.06445	0.08848	0.00028
2	1	0.80826	0.61766	0.50447	318	317	0.006057	6.93E-05	0.00161	9.73E-06	0.00027	0.00068	0.00081	0.02783	0.244	0.01376	0.0176	0.02698	0.04129	0.00028
3	0	0.81932	0.62347	0.54855	493	492	0.00391	5.99E-05	0.00075	2.93E-06	0.00025	0.00037	0.0567	0.512	0.02693	0.03141	0.0543	0.08037	0.00028	0.00028
3	0	0.80657	0.67256	0.61745	488	487	0.003956	5.38E-05	0.00083	3.79E-06	0.0001	0.00026	0.00029	0.06639	0.641	0.03747	0.05116	0.05412	0.11241	0.00028
3	0	0.82653	0.58326	0.44555	498	497	0.001873	3.34E-05	0.00069	2.68E-06	7.00E-05	0.00021	0.00022	0.05331	0.218	0.01281	0.0138	0.02246	0.03849	0.00028
4	0	0.8736	0.78996	0.78026	492	491	0.001924	6.72E-05	0.0021	1.18E-05	0.00077	0.00184	0.0023	0.38111	1.814	0.09816	0.14476	0.27649	0.24608	0.00028
4	0	0.81148	0.76831	0.70809	305	304	0.006316	0.000245	0.00341	2.16E-05	0.00093	0.00141	0.0028	0.13878	1.326	0.0722	0.0901	0.11271	0.21661	0.00028
4	0	0.80978	0.77992	0.69318	291	290	0.006424	0.002757	0.00457	3.03E-05	0.00159	0.00292	0.00477	0.11369	1.222	0.07043	0.09023	0.10485	0.21129	0.00028
5	1	0.81471	0.61483	0.53216	300	299	0.006433	3.86E-05	0.00085	5.45E-06	0.00017	0.00042	0.00051	0.04946	0.354	0.01756	0.02413	0.0469	0.05269	0.00028
5	1	0.83269	0.62018	0.57051	286	285	0.006754	5.17E-05	0.00111	7.52E-06	0.00024	0.00059	0.00072	0.02995	0.266	0.01211	0.01732	0.03501	0.03632	0.00028
5	1	0.82056	0.63244	0.57051	266	265	0.007257	4.86E-05	0.00086	6.28E-06	0.0002	0.00045	0.00059	0.02734	0.241	0.01289	0.01662	0.02548	0.03926	0.00028
6	1	0.79857	0.66865	0.44583	283	282	0.006824	0.000138	0.00177	1.11E-05	0.00025	0.00061	0.00075	0.1461	0.422	0.02602	0.02697	0.03944	0.07906	0.00028
6	1	0.79774	0.71199	0.36714	289	288	0.006893	6.49E-05	0.00122	8.13E-06	0.0002	0.00049	0.00061	0.08552	0.741	0.04596	0.05921	0.06201	0.13788	0.00028
6	1	0.82169	0.67061	0.36176	292	291	0.006424	2.80E-05	0.00084	5.58E-06	0.00018	0.00041	0.00055	0.02324	0.205	0.01087	0.01215	0.01213	0.03262	0.00028
7	1	0.43551	0.81029	0.71652	267	266	0.007227	0.0002	0.00063	4.72E-05	0.00244	0.00089	0.00731	0.22066	1.891	0.09811	0.10565	0.06615	0.29433	0.00028
7	1	0.7622	0.73507	0.75672	175	165	0.009677	0.001218	0.01268	0.000123	0.00494	0.00999	0.01482	0.13048	1.246	0.0612	0.06968	0.08788	0.18961	0.00028
7	1	0.18957	0.77921	0.66468	249	248	0.007184	9.15E-05	0.00111	2.13E-05	0.00085	0.00182	0.00235	0.057428	0.60095	0.01416	0.01131	0.05171	0.103148	0.00028
8	1	0.82878	0.67133	0.4662	262	261	0.007371	8.67E-05	0.00212	1.56E-05	0.00018	0.00093	0.00115	0.05143	0.315	0.0162	0.01983	0.03772	0.04859	0.00028
8	1	0.80489	0.62426	0.47446	285	284	0.006776	7.22E-05	0.0015	1.01E-05	0.00026	0.00068	0.00078	0.05138	0.293	0.01272	0.02001	0.03957	0.03816	0.00028
8	1	0.8227	0.62902	0.56263	289	288	0.006463	7.89E-05	0.00141	9.34E-06	0.00025	0.00062	0.00075	0.04543	0.405	0.01974	0.01997	0.04991	0.05021	0.00028
9	1	0.8621	0.60863	0.46038	332	331	0.005821	6.13E-05	0.00143	8.31E-06	0.00018	0.00049	0.00054	0.04009	0.348	0.02009	0.02365	0.03813	0.06026	0.00028
9	1	0.83607	0.61289	0.43513	333	332	0.005805	5.91E-05	0.00134	7.81E-06	0.00019	0.00051	0.00057	0.02599	0.497	0.0225	0.01388	0.05916	0.0675	0.00028
9	1	0.84942	0.60477	0.39889	331	330	0.005838	5.59E-05	0.0014	8.16E-06	0.0002	0.00053	0.0006	0.05496	0.3	0.01645	0.02051	0.03794	0.04934	0.00028
10	1	0.79198	0.66429	0.69379	212	211	0.009104	0.000111	0.00354	3.22E-05	0.00128	0.00228	0.00383	0.1214	1.105	0.05649	0.07947	0.12093	0.16946	0.00028
10	1	0.8396	0.64823	0.62499	248	247	0.007782	7.56E-05	0.00199	1.53E-05	0.00046	0.00138	0.00199	0.09149	0.817	0.03894	0.05824	0.10523	0.11862	0.00028
10	1	0.80884	0.64326	0.59772	237	236	0.00816	7.41E-05	0.00175	1.43E-05	0.00051	0.00105	0.00152	0.0944	0.898	0.03973	0.05641	0.10078	0.11919	0.00028
11	1	0.82127	0.64846	0.26212	375	374	0.005139	4.36E-05	0.00086	4.40E-06	9.00E-05	0.00025	0.00028	0.03564	0.316	0.02002	0.02042	0.01589	0.06005	0.00028
11	1	0.80379	0.64898	0.33536	292	291	0.006608	0.002308	0.00202	1.13E-05	0.00051	0.00065	0.00152	0.00027	0.526	0.01478	0.01486	0.04316	0.10433	0.00028
11	1	0.84385	0.67496	0.46658	285	283	0.006773	0.002461	0.00129	8.76E-06	0.00026	0.00058	0.00077	0.11166	1.002	0.06665	0.07022	0.08624	0.19695	0.00028
12	1	0.82004	0.74807	0.52028	260	259	0.007427	9.36E-05	0.00223	1.65E-05	0.00062	0.00122	0.00187	0.11983	1.047	0.07046	0.07085	0.08579	0.21137	0.00028
12	1	0.83261	0.71529	0.41361	265	264	0.007285	5.90E-05	0.002	1.46E-05	0.00057	0.00117	0.0017	0.04703	0.406	0.02621	0.02634	0.03051	0.07689	0.00028
12	1	0.82606	0.68839	0.47103	267	266	0.007219	9.42E-05	0.0018	1.30E-05	0.00018	0.00092	0.00115	0.06167	0.548	0.01105	0.03826	0.06166	0.0931	0.00028
13	1	0.81619	0.7231	0.32538	344	343	0.005614	6.24E-05	0.00074	4.18E-06	0.00013	0.00033	0.00039	0.02785	0.245	0.01574	0.01776	0.02094	0.04723	0.00028
13	1	0.80722	0.72248	0.46741	215	214	0.006018	0.002385	0.00229	2.64E-05	0.00063	0.00123	0.00178	0.0722	0.65	0.0351	0.04261	0.05007	0.1052	0.00028
13	1	0.80556	0.69601	0.49392	309	307	0.006239	0.001915	0.00129	8.04E-06	0.00034	0.00063	0.00101	0.05764	0.334	0.01886	0.02301	0.02666	0.05599	0.00028
14	0	0.86146	0.79945	0.57299	402	401	0.004809	3.95E-05	0.00096	4.61E-06	0.00023	0.00053	0.00068	0.05472	0.507	0.0319	0.03241	0.03717	0.09571	0.00028
14	0	0.88352	0.79509	0.36045	397	396	0.004861	3.36E-05	0.00145	6.97E-06	0.00028	0.00066	0.00081	0.04711	0.385	0.02033	0.02551	0.03452	0.06099	0.00028
14	0	0.86478	0.7657	0.37052	400	399	0.004827	5.38E-05	0.0014	6.75E-06	0.00022	0.00056	0.00065	0.03993	0.385	0.02053	0.02519	0.03499	0.06159	0.00028
15	1	0.79421	0.69206	0.59089	218	217	0.006107	0.000116	0.00205	1.44E-05	0.00051	0.0011	0.00153	0.09796	0.631	0.03007	0.04251	0.0725	0.09021	0.00028
15	1	0.7903	0.69206	0.67128	240	239	0.008054	0.000174	0.00224	1.80E-05	0.00064	0.00133	0.00193	0.09223	0.843	0.04111	0.05616	0.08603	0.12934	0.00028
15	1	0.73892	0.70522	0.70388	234	233	0.008258	0.000302	0.00396	2.94E-05	0.0004	0.002	0.00283	0.12446	1.127	0.0461	0.04723	0.13974	0.13829	0.00028
16	1	0.89951	0.72967	0.52362	314	313	0.004546	0.000217	0.002	0.00055	0.00061	0.00122	0.00146	0.1146	1.134	0.04648	0.05844	0.10384	0.19462	0.00028
16	0	0.7532	0.70885	0.6361	385	384	0.005016	0.001557	0.0036	1.80E-05	0.00087	0.0011	0.00261	0.15324	1.457	0.08481	0.10048	0.12695	0.25444	0.00028
16	0	0.80587	0.78348	0.7403	389	388	0.004801	0.000942	0.00314	1.51E-05	0.0009	0.0011	0.0027	0.20158	1.745	0.07798	0.13676	0.16788	0.29209	0.0002

# HOMEWORK

2. K-NN : we'll use fruits data as our example dataset.

<https://medium.com/analytics-vidhya/easy-knn-algorithm-using-scikit-learn-7f6e256c9453>



fruit\_data\_with\_colors.txt

# HOMEWORK

## 3. K mean Clustering : customer in mall – clustering with 3 features

<https://medium.com/data-folks-indonesia/step-by-step-to-understanding-k-means-clustering-and-implementation-with-sklearn-b55803f519d6>

Customer	Genre	Age	Annual Income (Spending \$)
1	Male	19	15
2	Male	21	15
3	Female	20	16
4	Female	23	16
5	Female	31	17
6	Female	22	17
7	Female	35	18
8	Female	23	18
9	Male	64	19
10	Female	30	19
11	Male	67	19
12	Female	35	19
13	Female	58	20
14	Female	24	20
15	Male	37	20
16	Male	22	20
17	Female	35	21
18	Male	20	21
19	Male	52	23
20	Female	35	23
21	Male	35	24
22	Male	25	24
23	Female	46	25
24	Male	31	25
25	Female	54	28
26	Male	29	28
27	Female	45	28
28	Male	35	28
29	Female	40	29
30	Female	23	29
31	Male	60	30
32	Female	21	30
33	Male	53	33
34	Male	18	33
35	Female	49	33
36	Female	21	33
37	Female	42	34
38	Female	30	34
39	Female	36	37
40	Female	20	37
41	Female	65	38
42	Male	24	38
43	Male	48	39
44	Female	31	39
45	Female	49	39
46	Female	24	39
47	Female	50	40
48	Female	27	40
49	Female	29	40
50	Female	31	40
51	Female	49	42
52	Male	33	42
53	Female	31	43
54	Male	59	43
55	Female	50	43
56	Male	47	43
57	Female	51	44
58	Male	60	44
59	Female	27	46
60	Male	53	46
61	Male	70	46
62	Male	19	46
63	Female	62	47
64	Female	54	47

Data: mall customer



# THE END