



Biological Sciences faculty
Biophysics Department



Classification Models in Machine Learning

Presented By
Alireza Doustmohammadi
Graduate Student in Bioinformatics

January 2021

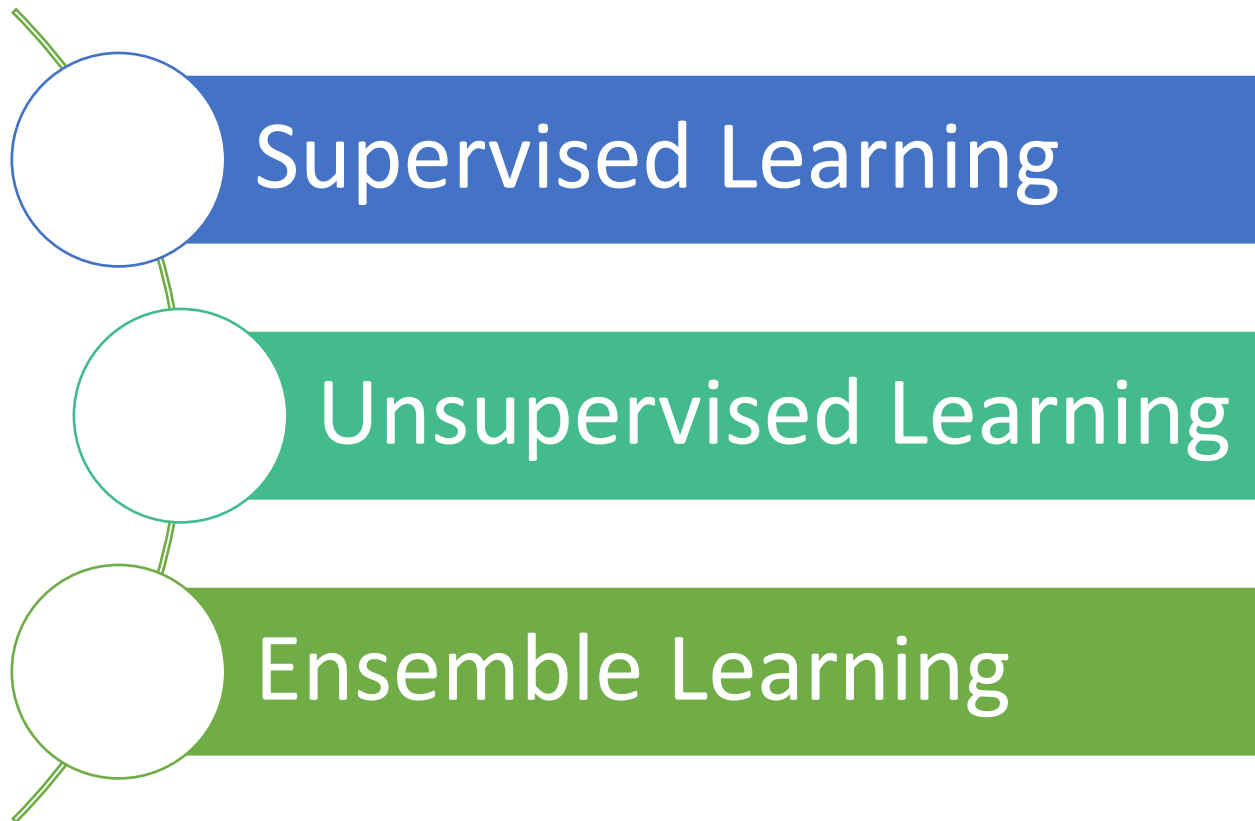
Contents

Introduction

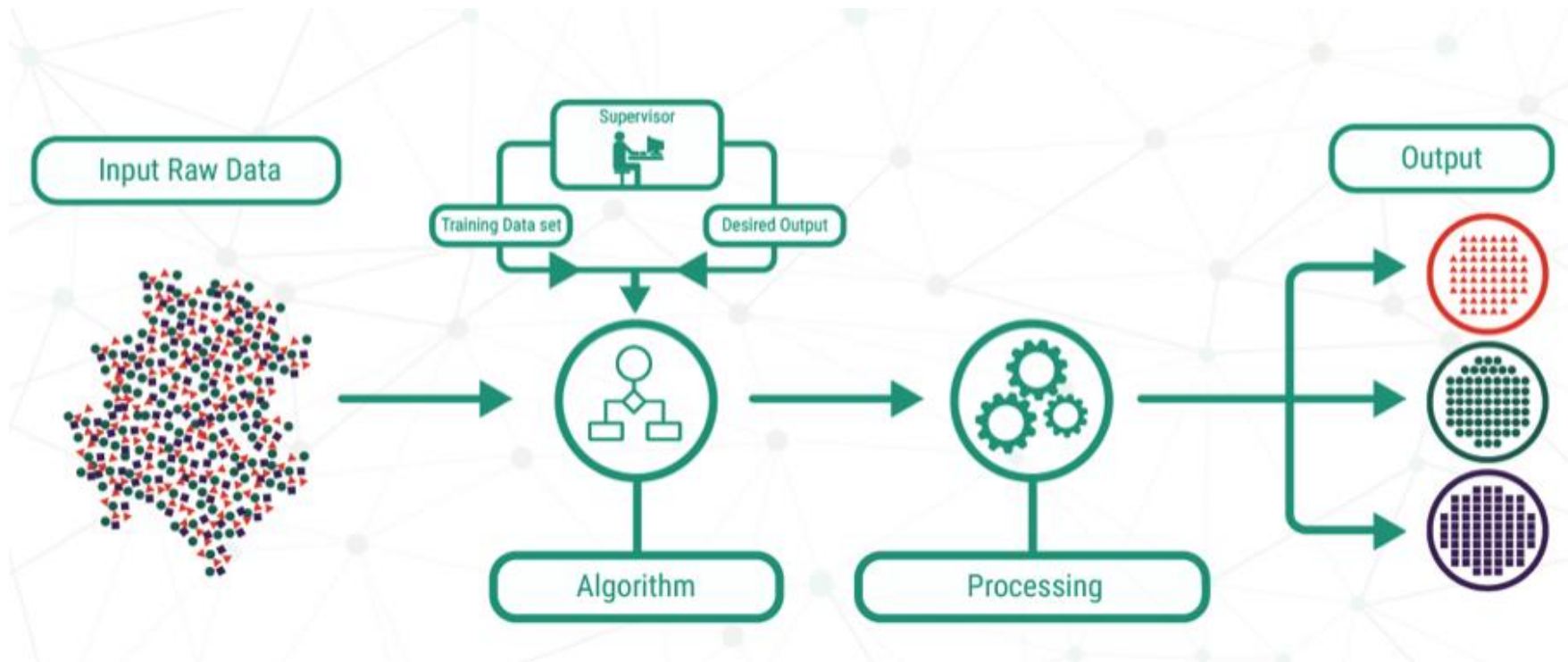
Supervised & Unsupervised Learning

Classification

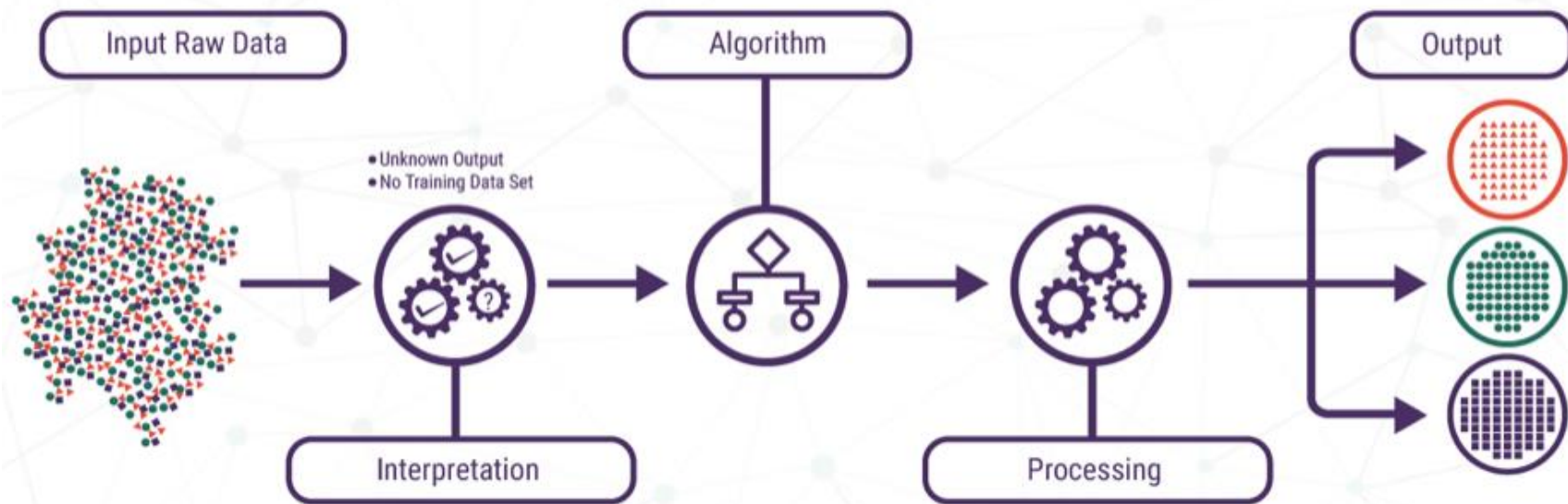
Outline



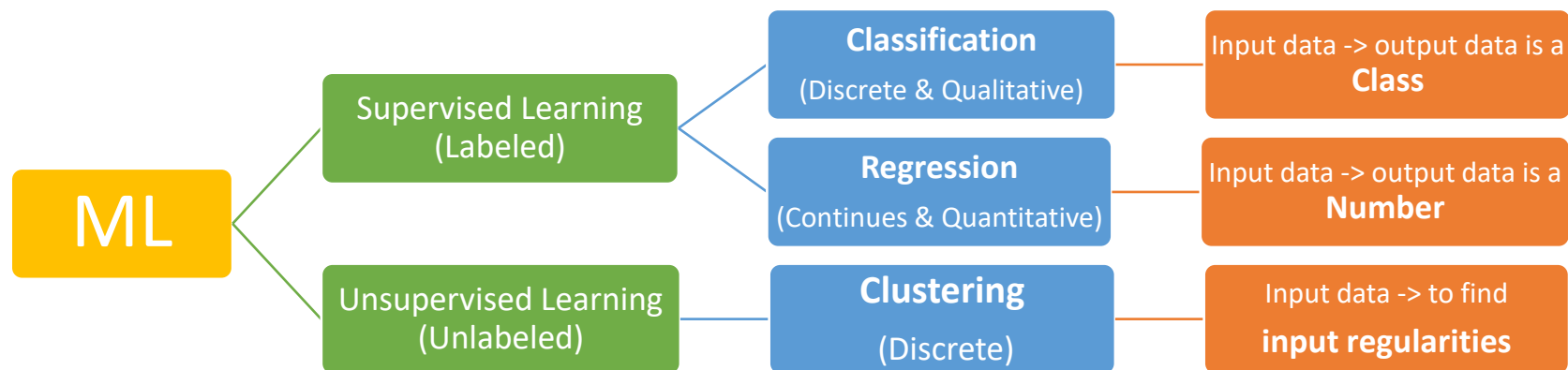
Supervised Learning



Unsupervised Learning



Supervised vs Unsupervised Learning



Classification





Naïve Bayes



Nearest Neighbor



Parzen Window

Naïve Bayes Classifier

❖ NB assign each observation to the most likely class

Probability of x occurring
given evidence c has already
occurred

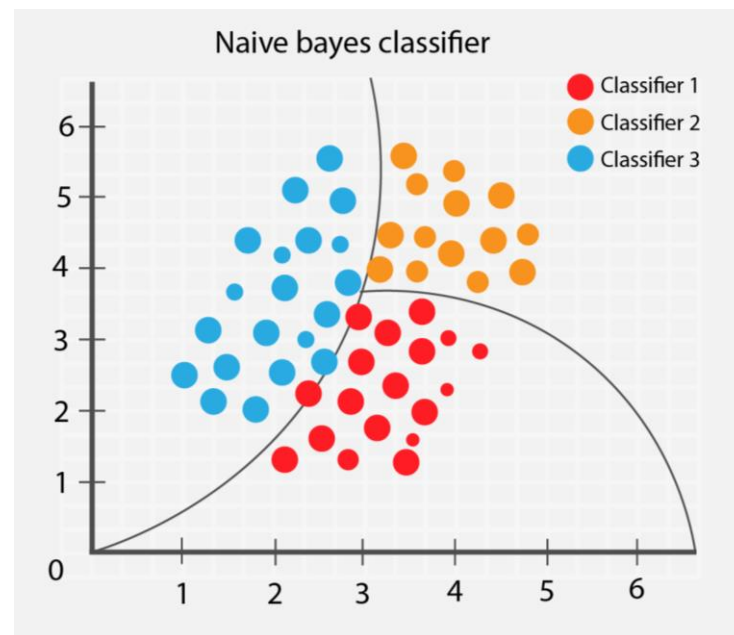
Class Prior Probability

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)}$$

Probability of c occurring
given evidence x has already
occurred

Predictor Prior Probability

$$\max Pr(Y = j|X)$$



Naïve Bayes Classifier

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

[https://www.saedsayad.com/naive_bayesian.htm]

Naïve Bayes Classifier

- Observation: sunny outlook

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \left\{ \begin{array}{l} P(NO|Sunny) = \frac{\frac{2}{5} \times \frac{5}{14}}{\frac{5}{14}} = \frac{2}{5} \\ P(YES|Sunny) = \frac{\frac{3}{9} \times \frac{9}{14}}{\frac{5}{14}} = \frac{3}{9} \end{array} \right. \rightarrow \max Pr(Y = j|x) = P(NO|Sunny)$$

Naïve Bayes Classifier

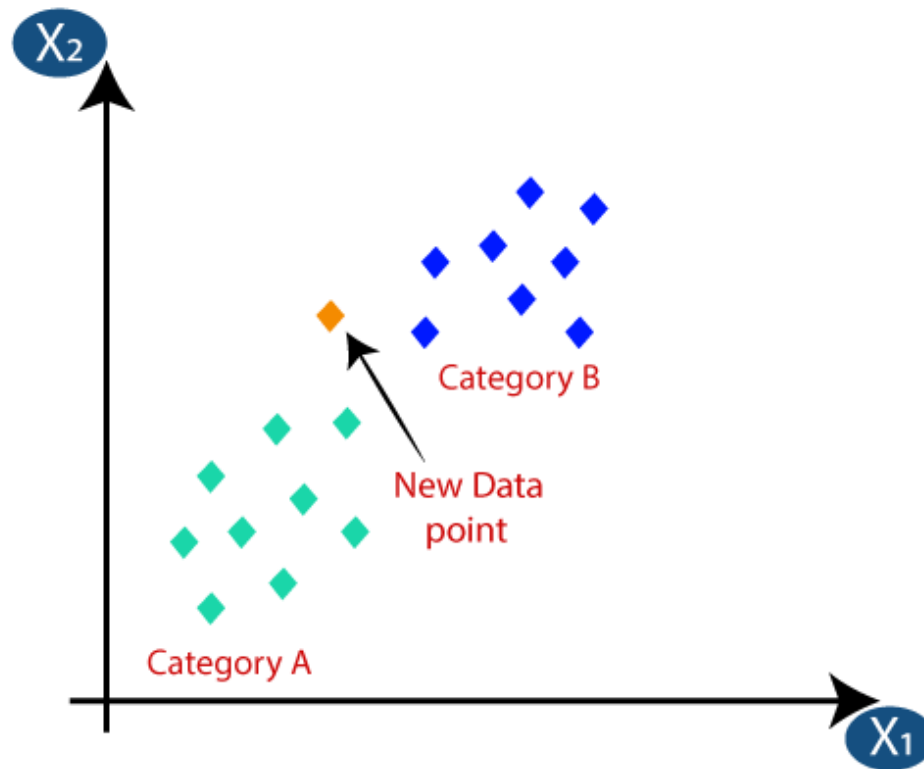
- **Observation: sunny outlook & Normal Humidity**

Frequency Table

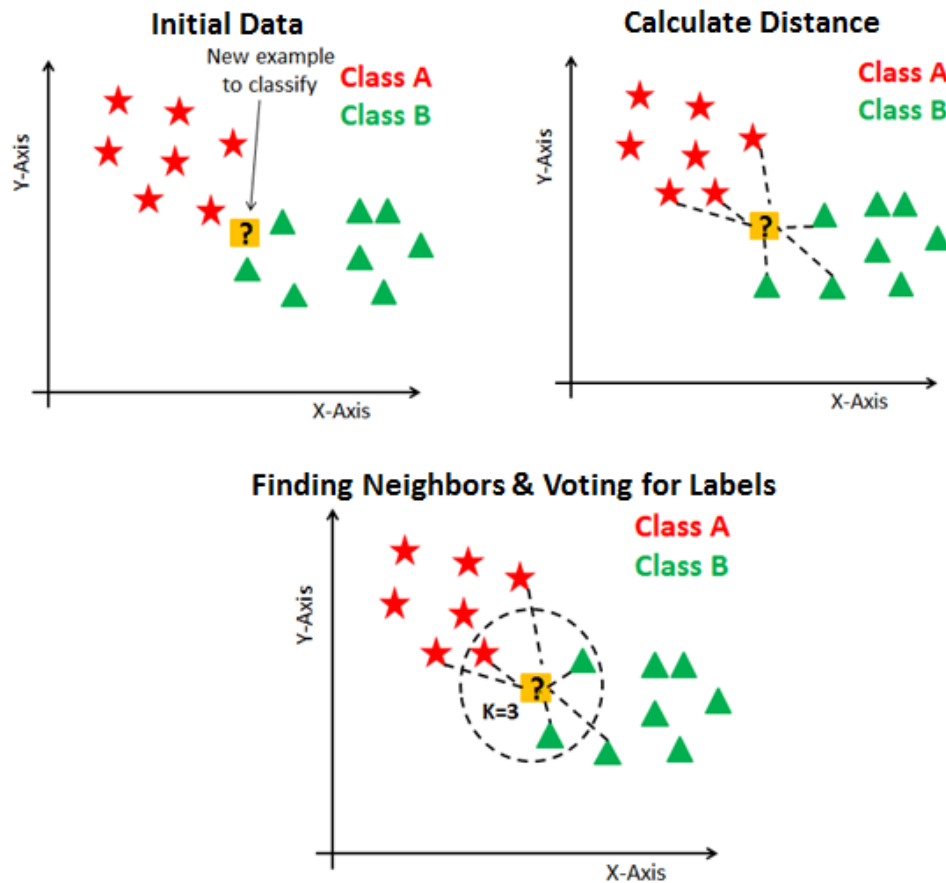
		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

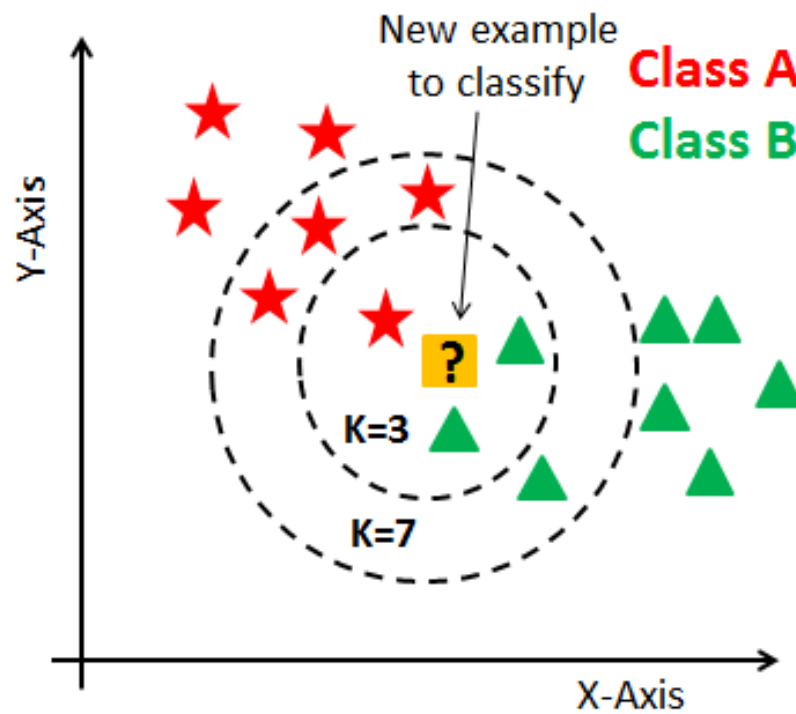
Nearest Neighbor Classifier



K Nearest Neighbor (KNN) Classifier



K Nearest Neighbor (KNN) Classifier



K Nearest Neighbor (KNN) Classifier

Distance Measures:

$$\text{Euclidean distance } (x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$\text{Manhattan distance } (x, y) = \sum_{i=1}^n |x_i - y_i|$$

$$\text{Camberra distance } (x, y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i + y_i|}$$

$$\text{Chebyshev distance } (x, y) = \max_{i=1}^n |x_i - y_i|$$

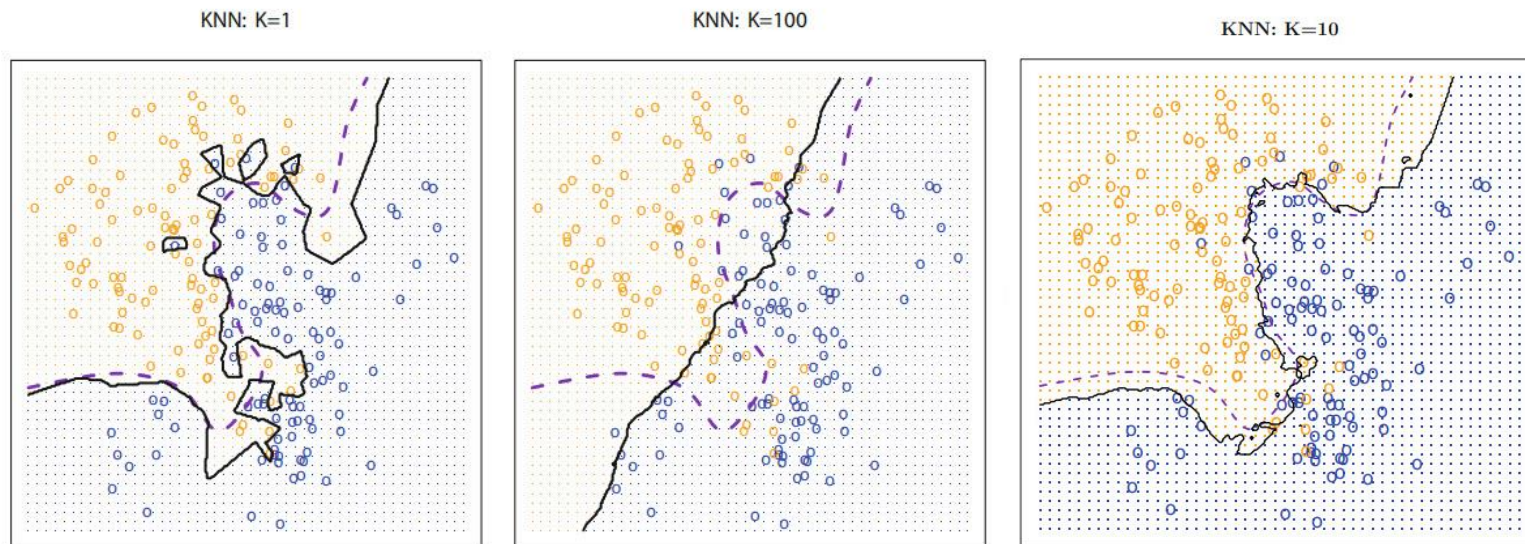
$$\text{Minkowski distance } (x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

K Nearest Neighbor (KNN) Classifier

z.average.s ynthesized	z.average.s erum	volume.m ean.synthe sized	volume.m ean.serum	number.m ean.synthe sized	number.m ean.serum	intensity.m ean.synthe sized	intensity.m ean.serum	polydisper sity.index.s ynthesized	polydisper sity.index.s erum	zeta.poten tial.synthe sized	zeta.poten tial.serum	lspr.index. synthesize d	lspr.index. serum	lspr.peak.p osition.syn thesized	serum.den sity.protot	serum.den sity.autot	total.spc	num.protei n.id	cell.associ ation
22.36	57.53	21.94	21.75	23.49	18.38	23.49	70.97	0.084	0.27	-21.78	-9.11	0.18253	0.454404	518.77	2.927	255.443	2028	94	Medium
30.95	90.06	11.76	67.79	47.5	53.87	47.5	106.7	0.399	0.215	15.22	-8.79	0.45821	0.525747	526.28	4.602	240.7288	2683	81	High
22.64	44.43	22.32	44.8	35.03	34.07	35.03	63.72	0.147	0.184	-24.08	-6.73	0.223534	0.274761	518.33	4.79	247.0191	1260	51	Medium
23.09	23.75	21.22	74.66	23.04	31.4	23.04	68.92	0.15	0.207	-20.27	-8.29	0.27362	0.327264	518.57	3.552	240.1767	1286	58	Medium
23.8	55.98	4.11	221.93	29.49	25.25	29.49	83.34	0.326	0.273	16.35	-7.95	0.365436	0.389573	523.23	5.741	244.2281	2401	81	High
25.22	38.8	21.32	36.99	23.12	34.38	23.12	41.33	0.144	0.154	-21.42	-6.54	0.20691	0.265327	519.87	2.976	259.0438	1557	71	Low
18.65	54.03	19.47	38.58	28.3	32.35	28.3	64.47	0.138	0.217	-25.26	-8.5	0.210431	0.292836	517.68	5.488	272.7686	1855	78	Medium
45.72	44.6	46.42	43.94	50.95	41.77	50.95	46.96	0.115	0.093	-12.29	-7.48	0.25094	0.280491	520.6	0.142	241.5201	1215	72	Low
19.85	38.69	20.14	33.52	21.49	30.34	21.49	40.44	0.086	0.154	-6.33	-5.91	0.211466	0.263698	519.07	1.67	258.6099	1717	91	Low
15.6	59.7	2.75	69.66	35.99	17.68	35.99	105.1	0.465	0.231	-5.01	-9.26	0.326142	0.365811	518.9	4.008	213.3235	1453	116	Medium
23.15	47.03	22.84	175.14	27.78	23.34	27.78	140.15	0.187	0.336	13.49	-9.7	0.266579	0.317134	520.23	0.3	234.6106	1098	75	Low
20.53	48.4	19.51	113.66	23.23	25.62	23.23	99.43	0.191	0.308	19.6	-6.52	0.275461	0.324751	519.23	1.215	247.7433	1431	87	Low
28.17	47.34	28.17	62.91	31.58	34.51	31.58	83.76	0.193	0.223	13.92	-6.23	0.276498	0.297223	522.1	2.107	235.4843	1577	73	High
33.6	58.95	16.66	33.85	28.99	26.12	28.99	87.33	0.268	0.346	25.83	-5.46	0.309989	0.367331	521.43	0.412	257.8317	231	28	Medium
82.41	59.93	27.08	33.04	55.31	27.7	55.31	86.81	0.249	0.367	-29.23	-6.1	0.395907	0.320779	534.13	0.526	249.1523	1706	53	Medium
27.94	100.13	16.05	124.61	35.19	29.7	35.19	142.63	0.302	0.23	-21.68	-5.71	0.465011	0.359906	550.6	0.463	208.3178	1366	85	Low
23.58	60.27	22.67	45.24	25.16	32.93	25.16	81.13	0.108	0.248	-24.32	-10.42	0.241281	0.413864	522.73	5.883	271.1633	1723	83	Medium
42.05	46.63	31.55	57	36.47	38.14	36.47	72.78	0.216	0.197	-8.96	-6.63	0.175809	0.244867	519.23	2.921	238.2519	2295	104	Medium
77.02	55.39	26.46	45.34	38.3	35.75	38.3	74.16	0.211	0.236	-17.77	-5.68	0.261762	0.402932	518.6	4.049	243.9545	1321	56	Medium

K Nearest Neighbor (KNN) Classifier

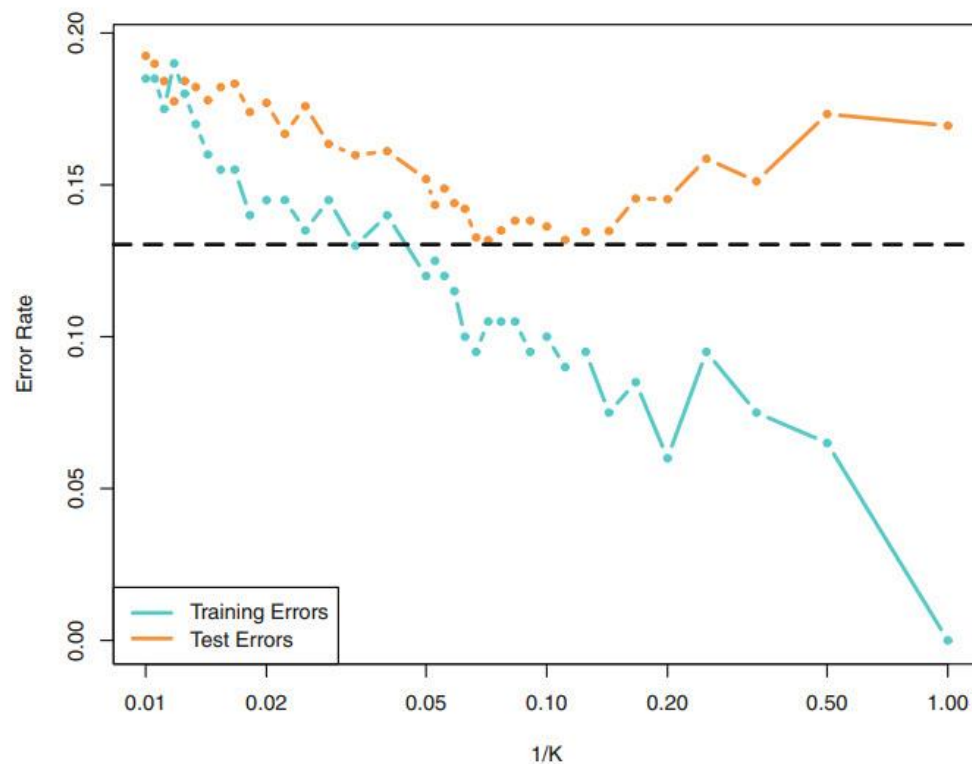
Number of Neighbors:



✓ Selecting Appropriate k

K Nearest Neighbor (KNN) Classifier

Relation between k and *Flexibility*:



Weighted K Nearest Neighbor (WKNN) Classifier

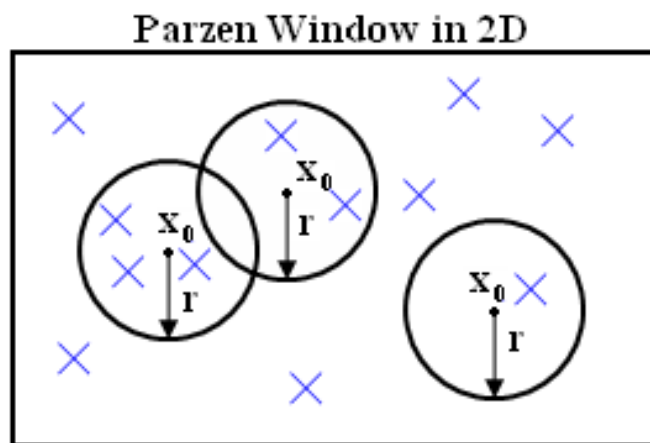
❖ Distance-weighted voting

Closer neighbors receive a higher weight
and more influence to derive the output value

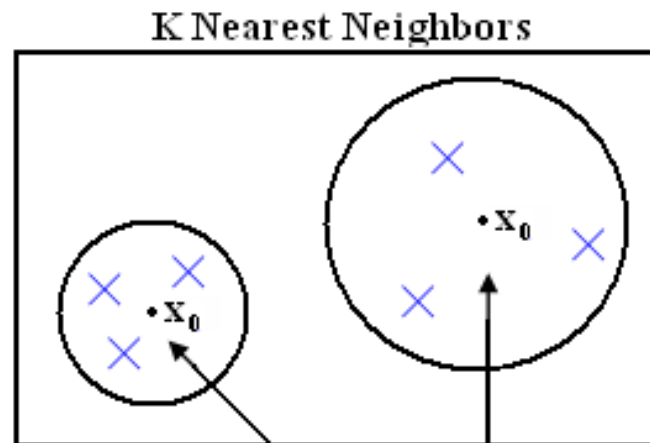
K Nearest Neighbor (KNN) Classifier

- ❖ Lazy Learning
- ❖ Instance-based Learning

Parzen Window Classifier



same r for any x_0



contain K sample points

“Machine learning is the next internet”

-Anthony Tether

Director, DARPA (Defense Advanced Research
Projects Agency, USA).