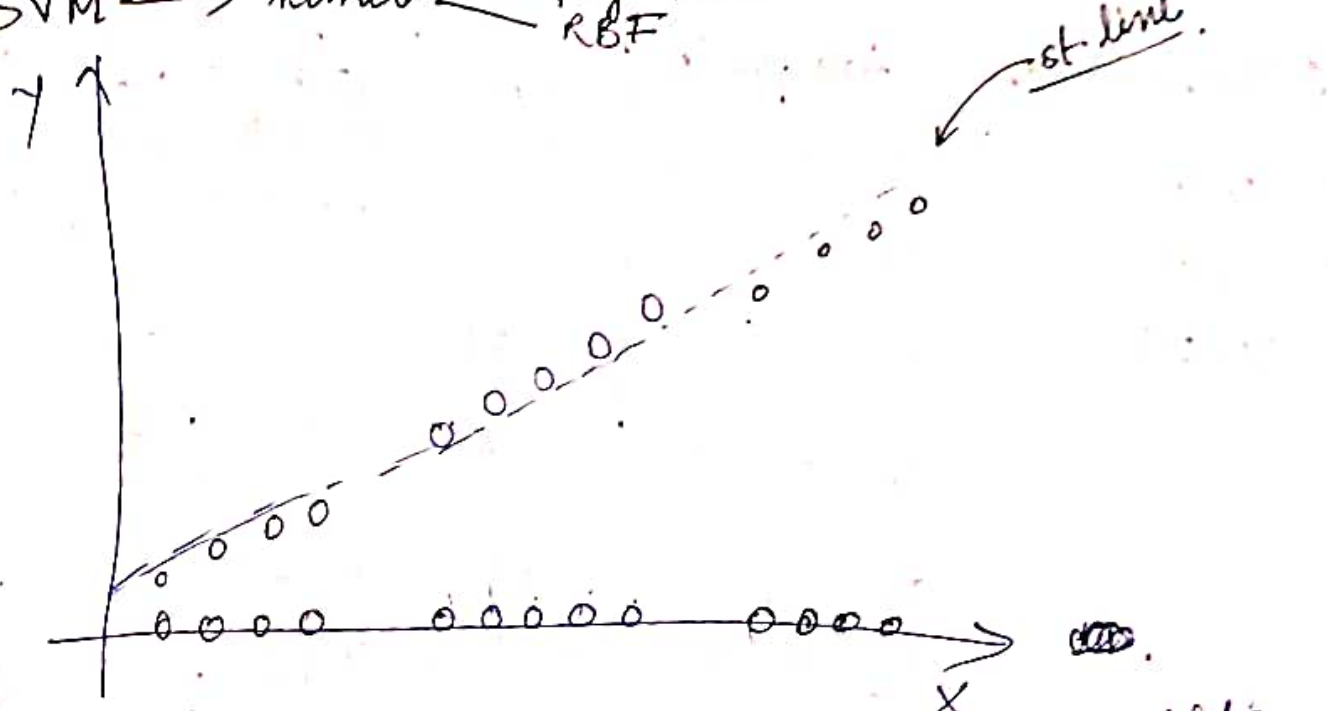
21.09.24

SVM → kernel → linear
polynomial
RBF



x = weight
y = (weight)²
plot (w, w²).

- Transform the data points to diff dimension using kernel so as it is easy to classify the scattered datapoints.

27.09.24

Unsupervised learning → K means clustering

→ Group of documents are categorized to similar sets of cluster.

Supervised classification — in this label is given

	Eg:		label/class	
		d_1	A	Training is done.
		d_2	A	
		d_3	B	
		d_n	C	

Apply learning → d_{new} → classify the new data.

27.09.24 Unsupervised. 1st iteration.

Initial centroid

1: (2,10)

2: (5,8)

3: (1,2)

Data points			Distance to 1 (2,10)	2 (5,8)	3 (1,2)	ob
D_1	2	10	0	3.61	8.06	
D_2	2	5	5	4.24	3.16	
D_3	8	4	8.48	5	7.28	
D_4	5	8	3.61	0	7.21	
D_5	7	5	7.07	3.61	6.71	2
D_6	6	4	7.21	4.12	5.39	2
D_7	1	2	8.06	7.21	0	3
D_8	4	9	2.24	1.41	7.62	2

Euclidian distance $= \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

New centroid: (Cluster – data points mean).

1: (2,10)

2: $\left(\frac{8+5+7+6+4}{5}, \frac{4+8+5+4+9}{5}\right)$

$= (6, 6)$

3: $\left(\frac{2+1}{2}, \frac{5+2}{2}\right)$

$= (1.5, 3.5)$

2nd iteration 27.09.24

Datapoints		1 (2,10)	2 (6,6)	3 (1.5,3.5)	Cluster
2	10	0	5.66	6.52	1
2	5	5	4.12	1.58	3
8	4	8.49	2.83	6.52	2
5	8	3.61	2.824	5.70	2
7	5	7.07	1.41	5.70	2
6	4	7.21	2	4.53	2
1	2	8.06	6.40	1.58	3
4	9	2.24	3.61	6.04	1

Not same cluster

New centroid.

1: $\left(\frac{2+4}{2}, \frac{10+9}{2}\right)$

$= (3, 9.5)$

2: $\left(\frac{8+5+7+6}{4}, \frac{4+8+5+4}{4}\right)$

$= (6.5, 5.25)$

3: $(1.5, 3.5)$

Algo:

1. identify centroid
2. calc dist.
3. Assign cluster

↓

Repeat until converge.

3rd iteration 27.09.24

Data point		1 (3, 9.5)	2 (6.5, 5.25)	3 (1.5, 3.5)	cluster
2	10	1.12	6.54	6.52	1
2	5	4.61	4.51	1.58	3
8	4	7.43	1.95	6.52	2
5	8	2.5	3.13	5.7	1
7	5	6.02	0.56	5.7	2
6	4	6.26	1.35	4.53	2
1	2	7.76	6.39	1.58	3
4	9	1.12	4.51	6.04	1

Cont. till convergence.

28.09.24

8.

Given 15 data points

$D_1(2,10)$ $D_2(2,6)$ $D_3(11,11)$ $D_4(6,9)$ $D_5(6,4)$

$D_6(1,2)$ $D_7(5,10)$ $D_8(4,9)$ $D_9(10,12)$ $D_{10}(7,5)$

$D_{11}(9,11)$ $D_{12}(4,6)$ $D_{13}(8,10)$ $D_{14}(3,8)$ $D_{15}(3,16)$

3 initial centroid

C1: (2, 6)
C2: (5, 10)
C3: (6, 11)

Using K means clustering - Iteration 1 28.09.24

Data points.		(2,6)	(5,10)	(6,11)	Cluster
2	10	4	3	5	2
2	6	0	7	9	1
11	11	14	7	5	3
6	9	7	2	2	2
6	4	6	7	7	1
1	2	5	12	14	1
5	10	7	0	2	2
4	9	5	2	4	2
10	12	14	7	5	3
7	5	6	7	7	1
9	11	12	5	3	3
4	6	2	5	7	1
3	10	5	02	4	2
3	8	3	4	6	1
3	16.	11	8	8	2

C1: $\left(\frac{2+6+1+7+4+3}{6}, \frac{6+4+2+5+6+8}{6}\right)$

$\left(\frac{23}{6}, \frac{31}{6}\right)$

$(3.83, 5.167)$

C2: $\left(\frac{2+6+5+4+3+3}{6}, \frac{10+9+10+9+10+16}{6}\right)$

$\left(\frac{23}{6}, \frac{64}{6}\right) = (3.83, 10.67)$

28.09.24

C3: $\left(\left(\frac{11+10+9}{3}\right), \left(\frac{11+12+11}{3}\right)\right)$

$= \left(\frac{30}{3}, \frac{34}{3}\right)$

$= (10, 11.33)$.

Iteration 2

Datapts	1 (3.83, 5.17)	2 (3.83, 10.67)	3 (10, 11.33)	cl
2 10	8.66	2.5	9.33	
2 6	2.66	6.5	12.33	
11 11	13	7.5	1.33	
6 9	6	3.84	6.33	
6 4	3.34	8.84	11.33	
1 2	6	11.5	18.33	
5 10	6	1.84	6.33	
4 9	4	1.84.	8.33	
10 12	13	7.5	0.67	
7 5	3.34	8.84	9.33	
9 11	11	5.5	1.33	
4 6	1	4.84	11.33	
3 10	5.66	1.5	8.33	
3 8	3.66	3.5	10.33	
3 16	11.66	6 16	11.67	

28.09.24

C1: $\left(\frac{20}{5}, \frac{23}{5}\right) = (4, 4.6)$

C2: $\left(\frac{2+1+5+3+13+13}{6}, \frac{10+12+10+10+8+16}{6}\right)$

$= (17/6, 56/6) = (2.83, 9.33)$

C3: (10, 11; 53)

Heirarichal clustering

1. given n data pts.
2. calc. dist. b/w every data pt & store in a matrix
3. identify the least dist. data pts (2 nearest data pts.) & merge them
4. Repeat step 3 untill all data pts. comes under one single cluster.

	A	B	C	D
A	0	1	4	5
B		0	2	6.1.5
C			0	3.1.5
D				0