Jeremy Yang

ADS Assignment 8

1. K-Means: 2D points $(x,y) \in \underset{1}{(2,5)}, \underset{2}{(1,5)}, \underset{3}{(22,55)}, \underset{4}{(42,12)}, \underset{5}{(15,16)}$ (i)

Cluster for $K=2$

Step 1: Init centroids randomly: $\mu_1 = (17, 43)$ $\mu_2 = (36, 6)$

Iteration 1:

a) Assign cluster by closest centroid

$$c^{(1)} = \arg\min_j (15^2+38^2, 34^2+1^2) = 2$$
$$c^{(2)} = \arg\min_j (16^2+38^2, 35^2+1^2) = 2$$
$$c^{(3)} = \arg\min_j (5^2+12^2, 14^2+49^2) = 1$$
$$c^{(4)} = \arg\min_j (25^2+31^2, 6^2+6^2) = 2$$
$$c^{(5)} = \arg\min_j (\overset{733}{2^2+27^2}, \overset{551}{21^2+10^2}) = 2$$

b) Move centroids

$$\mu_1 = (22, 55)$$
$$\mu_2 = \left(\frac{2+1+42+15}{4}, \frac{5+5+12+16}{4}\right) = (15, 9.5)$$

Iteration 2:

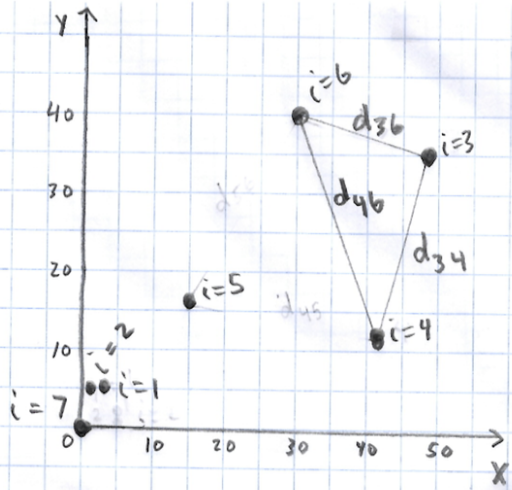
$$c^{(1)} = \arg\min_j (20^2+50^2, 13^2+4.5^2) = 2$$
$$c^{(2)} = \arg\min_j (21^2+50^2, 14^2+4.5^2) = 2$$
$$c^{(3)} = \arg\min_j (0+0, 40^2+45.5^2) = 1$$
$$c^{(4)} = \arg\min_j (20^2+43^2, 27^2+2.5^2) = 2$$
$$c^{(5)} = \arg\min_j (7^2+39^2, 0+6.5^2) = 2$$

converged!

ADS Assignment 8

K-Nearest Neighbors

	given				determine			
(x_i, y_i)	(2,5)	(1,5)	(48,35)	(42,12)	(15,16)	(30,40)	(0,0)	
class	1	1	2	2	1	2	1	K=1
	1	1	2	2	1	2	1	K=2
	1	1	2	2	1	2	1	K=3
i	1	2	3	4	5	6	7	



compute distances not obvious for ranking nearness:

$$d_{34} = \sqrt{6^2 + 23^2} = 23.8$$

ADS Assignment 8 JJ Yang

Decision Trees

1.

Entropy = un-homogeneity within groupings

$$= -\sum_{k=1}^{K} \hat{p}_{mk} \log(\hat{p}_{mk})$$

where $\hat{p}$ = sample prob

m = grouping

k = outcome/class

Information gain = reduction of entropy = $E_n - E_{n-1}$

= reduction of mis-classification error

2.

A	B	C	Label
a	a	a	1
b	b	a	2
a	a	b	1
b	b	a	2

$E_{initial}$

$= -.5 \ln(.5) = .69$

Feature A split:

$$E_A = -\sum_{k=1}^{2} \hat{p}_{mk} \log(\hat{p}_{mk}) = -1 \ln(1) - 1 \ln(1) = 0$$

Feature B split:

$$E_B = \quad -1 \ln(1) - 1 \ln(1) = 0$$

Feature C split:

$$E_c = \quad -.5 \ln(.5) - .5 \ln(.5) = .69$$

ADS Assignment 9

JJ Yang

Naïve Bayes

Temp	Windy	Humid	Rain
H	1	L	0
H	1	M	1
L	1	H	1
L	1	L	0
H	0	L	0
H	0	M	1
L	0	L	0
L	0	M	0

$$P(R \mid T{=}H \,\&\, W \,\&\, H{=}H) = P(T{=}H \,\&\, W \,\&\, H{=}H \mid R) \cdot P(R) \quad \text{(Bayes)}$$

$$= P(T{=}H \mid R) \cdot P(W \mid R) \cdot P(H{=}H \mid R) \cdot P(R) \quad \text{(independence)}$$

$$= .67 \cdot .67 \cdot .33 \cdot .375 = \underline{.0556}$$

$$P(!R \mid T{=}H \,\&\, W \,\&\, H{=}H) = P(T{=}H \mid !R) \cdot P(W \mid !R) \cdot P(H{=}H \mid !R) \cdot P(!R)$$

$$= .25 \cdot .4 \cdot 0 \cdot .625 = \underline{0}$$

From data P rain = 1 !