

Homework 2

Question 1a

$$\begin{aligned} Gini(price) &= \frac{4}{11} \left(1 - \left(\frac{2}{4} \right)^2 - \left(\frac{2}{4} \right)^2 \right) + \frac{4}{11} \left(1 - \left(\frac{2}{4} \right)^2 - \left(\frac{2}{4} \right)^2 \right) + \frac{3}{11} \left(1 - \left(\frac{2}{3} \right)^2 - \left(\frac{1}{3} \right)^2 \right) \\ &= 0.4848 \end{aligned}$$

$$\begin{aligned} Gini(maintenance) &= \frac{2}{11} \left(1 - \left(\frac{2}{2} \right)^2 - \left(\frac{0}{2} \right)^2 \right) + \frac{4}{11} \left(1 - \left(\frac{2}{4} \right)^2 - \left(\frac{2}{4} \right)^2 \right) + \frac{5}{11} \left(1 - \left(\frac{2}{5} \right)^2 - \left(\frac{3}{5} \right)^2 \right) \\ &= 0.4 \end{aligned}$$

$$\begin{aligned} Gini(capacity) &= \frac{3}{11} \left(1 - \left(\frac{1}{3} \right)^2 - \left(\frac{2}{3} \right)^2 \right) + \frac{6}{11} \left(1 - \left(\frac{3}{6} \right)^2 - \left(\frac{3}{6} \right)^2 \right) + \frac{2}{11} \left(1 - \left(\frac{0}{2} \right)^2 - \left(\frac{2}{2} \right)^2 \right) \\ &= 0.3939 \end{aligned}$$

$$\begin{aligned} Gini(airbag) &= \frac{5}{11} \left(1 - \left(\frac{3}{5} \right)^2 - \left(\frac{2}{5} \right)^2 \right) + \frac{6}{11} \left(1 - \left(\frac{3}{6} \right)^2 - \left(\frac{3}{6} \right)^2 \right) \\ &= 0.4909 \end{aligned}$$

Selected root: *capacity*

Question 1b

$$\begin{aligned} Entropy((1)price) &= \frac{8}{11} \left(-\frac{4}{8} \log_2 \frac{4}{8} - \frac{4}{8} \log_2 \frac{4}{8} \right) + \frac{3}{11} \left(-\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \right) \\ &= 0.9777 \end{aligned}$$

$$\begin{aligned} Entropy((2)maintenance) &= \frac{5}{11} \left(-\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \right) + \frac{6}{11} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) \\ &= 0.9422 \end{aligned}$$

$$\begin{aligned} Entropy((3)maintenance) &= \frac{9}{11} \left(-\frac{4}{9} \log_2 \frac{4}{9} - \frac{5}{9} \log_2 \frac{5}{9} \right) + \frac{2}{11} \left(-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} \right) \\ &= 0.8109 \end{aligned}$$

$$\begin{aligned} Entropy((4)capacity) &= \frac{3}{11} \left(-\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \right) + \frac{8}{11} \left(-\frac{5}{8} \log_2 \frac{5}{8} - \frac{3}{8} \log_2 \frac{3}{8} \right) \\ &= 0.9446 \end{aligned}$$

Best root: (3)*maintenance*

Question 2

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	5	6.082763	4.242641	5	4.123106	3.162278	4.472136
A2	5	0	8.485281	3.605551	7.071068	7.211103	8.062258	2.236068
A3	6.082763	8.485281	0	5	1.414214	2	7.28011	6.403124
A4	4.242641	3.605551	5	0	3.605551	4.123106	7.211103	1.414214
A5	5	7.071068	1.414214	3.605551	0	1.414214	6.708204	5
A6	4.123106	7.211103	2	4.123106	1.414214	0	5.385165	5.385165
A7	3.162278	8.062258	7.28011	7.211103	6.708204	5.385165	0	7.615773
A8	4.472136	2.236068	6.403124	1.414214	5	5.385165	7.615773	0

Single-link

	A1	A2	(A3, A5)	A4	A6	A7	A8
A1	0	5	5	4.242641	4.123106	3.162278	4.472136
A2		0	7.071068	3.605551	7.211103	8.062258	2.236068
(A3, A5)			0	3.605551	1.414214	6.708204	5
A4				0	4.123106	7.211103	1.414214
A6					0	5.385165	5.385165
A7						0	7.615773
A8							0

	A1	A2	((A3, A5), A6)	A4	A7	A8
A1	0	5	4.123106	4.242641	3.162278	4.472136
A2		0	7.071068	3.605551	8.062258	2.236068
((A3, A5), A6)			0	3.605551	5.385165	5
A4				0	7.211103	1.414214
A7					0	7.615773
A8						0

	A1	A2	((A3, A5), A6)	(A4, A8)	A7
A1	0	5	4.123106	4.242641	3.162278
A2		0	7.071068	2.236068	8.062258
((A3, A5), A6)			0	3.605551	5.385165
(A4, A8)				0	7.211103
A7					0

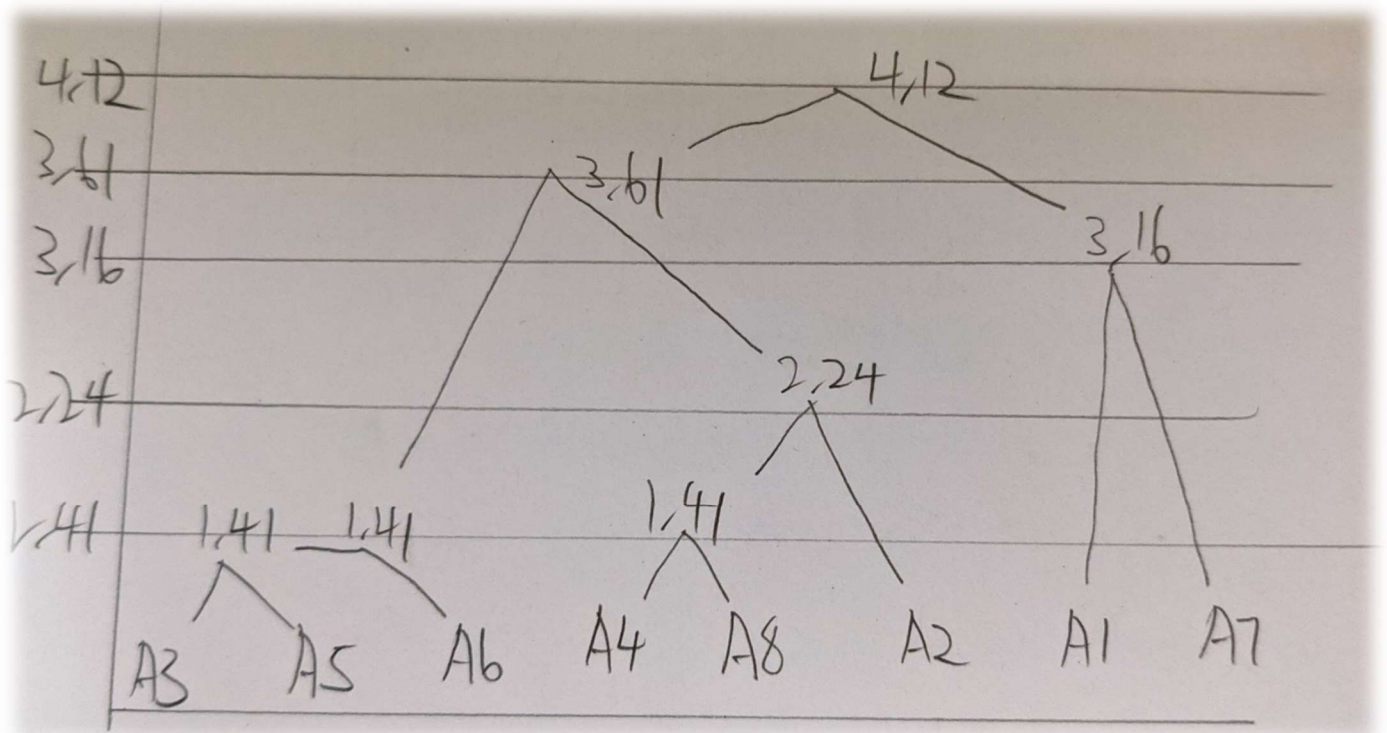
	A1	(A2, (A4, A8))	((A3, A5), A6)	A7
A1	0	4.242641	4.123106	3.162278
(A2, (A4, A8))		0	3.605551	7.211103
((A3, A5), A6)			0	5.385165
A7				0

	(A1, A7)	(A2, (A4, A8))	((A3, A5), A6)
(A1, A7)	0	4.242641	4.123106
(A2, (A4, A8))		0	3.605551
((A3, A5), A6)			0

	(A1, A7)	((A2, (A4, A8)), ((A3, A5), A6))
(A1, A7)	0	4.123106
((A2, (A4, A8)), ((A3, A5), A6))		0

	((A1, A7), ((A2, (A4, A8)), ((A3, A5), A6)))
((A1, A7), ((A2, (A4, A8)), ((A3, A5), A6)))	0

Dendrograms



Complete-link

	A1	A2	(A3, A5)	A4	A6	A7	A8
A1	0	5	6.082763	4.242641	4.123106	3.162278	4.472136
A2		0	8.485281	3.605551	7.211103	8.062258	2.236068
(A3, A5)			0	5	2	7.28011	6.403124
A4				0	4.123106	7.211103	1.414214
A6					0	5.385165	5.385165
A7						0	7.615773
A8							0

	A1	A2	(A3, A5)	(A4, A8)	A6	A7
A1	0	5	6.082763	4.472136	4.123106	3.162278
A2		0	8.485281	3.605551	7.211103	8.062258
(A3, A5)			0	6.403124	2	7.28011
(A4, A8)				0	5.385165	7.615773
A6					0	5.385165
A7						0

	A1	A2	((A3, A5), A6)	(A4, A8)	A7
A1	0	5	6.082763	4.472136	3.162278
A2		0	8.485281	3.605551	8.062258
((A3, A5), A6)			0	6.403124	7.28011
(A4, A8)				0	7.615773
A7					0

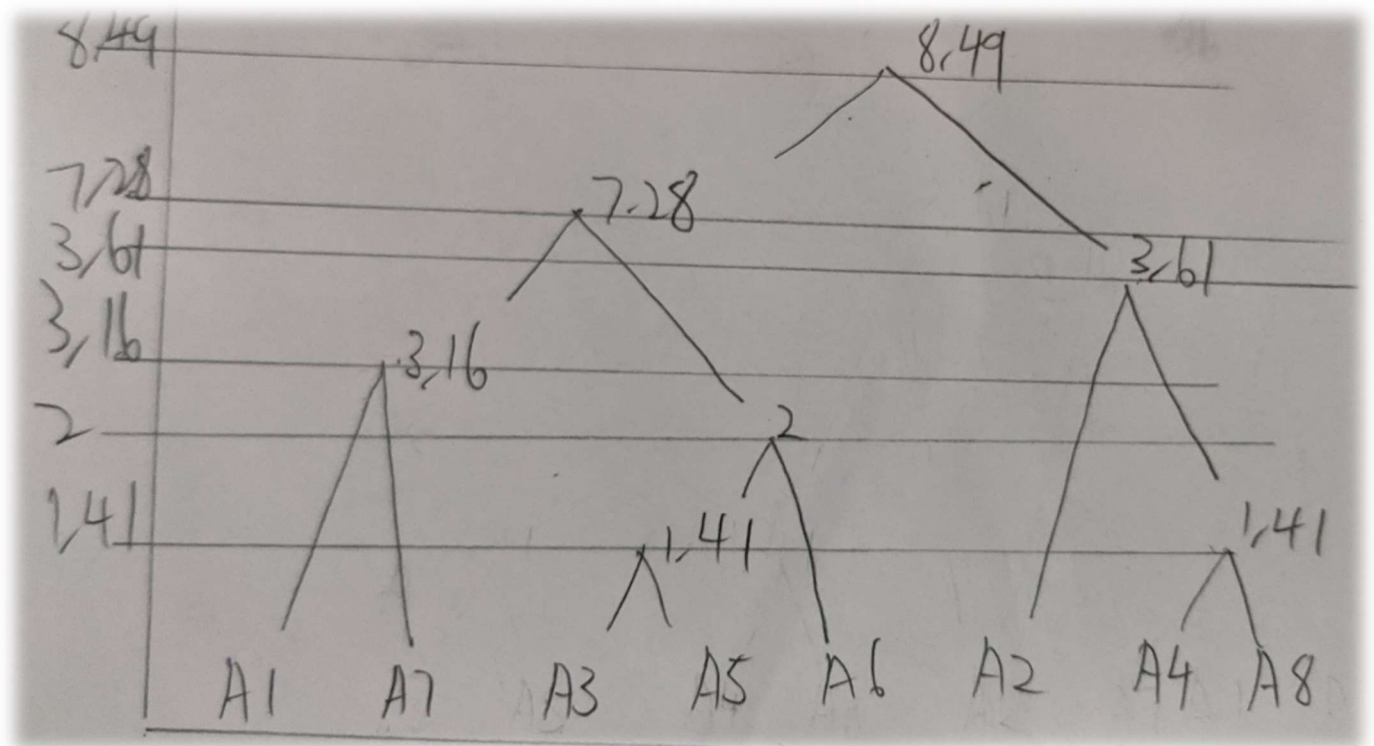
	(A1, A7)	A2	((A3, A5), A6)	(A4, A8)
(A1, A7)	0	8.062258	7.28011	7.615773
A2		0	8.485281	3.605551
((A3, A5), A6)			0	6.403124
(A4, A8)				0

	(A1, A7)	(A2, (A4, A8))	((A3, A5), A6)
(A1, A7)	0	8.062258	7.28011
(A2, (A4, A8))		0	8.485281
((A3, A5), A6)			0

	((A1, A7), ((A3, A5), A6))	(A2, (A4, A8))
((A1, A7), ((A3, A5), A6))	0	8.485281
(A2, (A4, A8))		0

	((((A1, A7), ((A3, A5), A6)), (A2, (A4, A8))))
((((A1, A7), ((A3, A5), A6)), (A2, (A4, A8))))	0

Dendrograms



Average-link

	A1	A2	(A3, A5)	A4	A6	A7	A8
A1	0	5	5.541382	4.242641	4.123106	3.162278	4.472136
A2		0	7.778175	3.605551	7.211103	8.062258	2.236068
(A3, A5)			0	4.302776	1.707107	6.994157	5.701562
A4				0	4.123106	7.211103	1.414214
A6					0	5.385165	5.385165
A7						0	7.615773
A8							0

	A1	A2	(A3, A5)	(A4, A8)	A6	A7
A1	0	5	5.541382	4.357389	4.123106	3.162278
A2		0	7.778175	2.920810	7.211103	8.062258
(A3, A5)			0	5.002169	1.707107	6.994157
(A4, A8)				0	4.754136	7.413438
A6					0	5.385165
A7						0

	A1	A2	((A3, A5), A6)	(A4, A8)	A7
A1	0	5	4.832244	4.357389	3.162278
A2		0	7.494639	2.920810	8.062258
((A3, A5), A6)			0	4.878153	6.189661
(A4, A8)				0	7.413438
A7					0

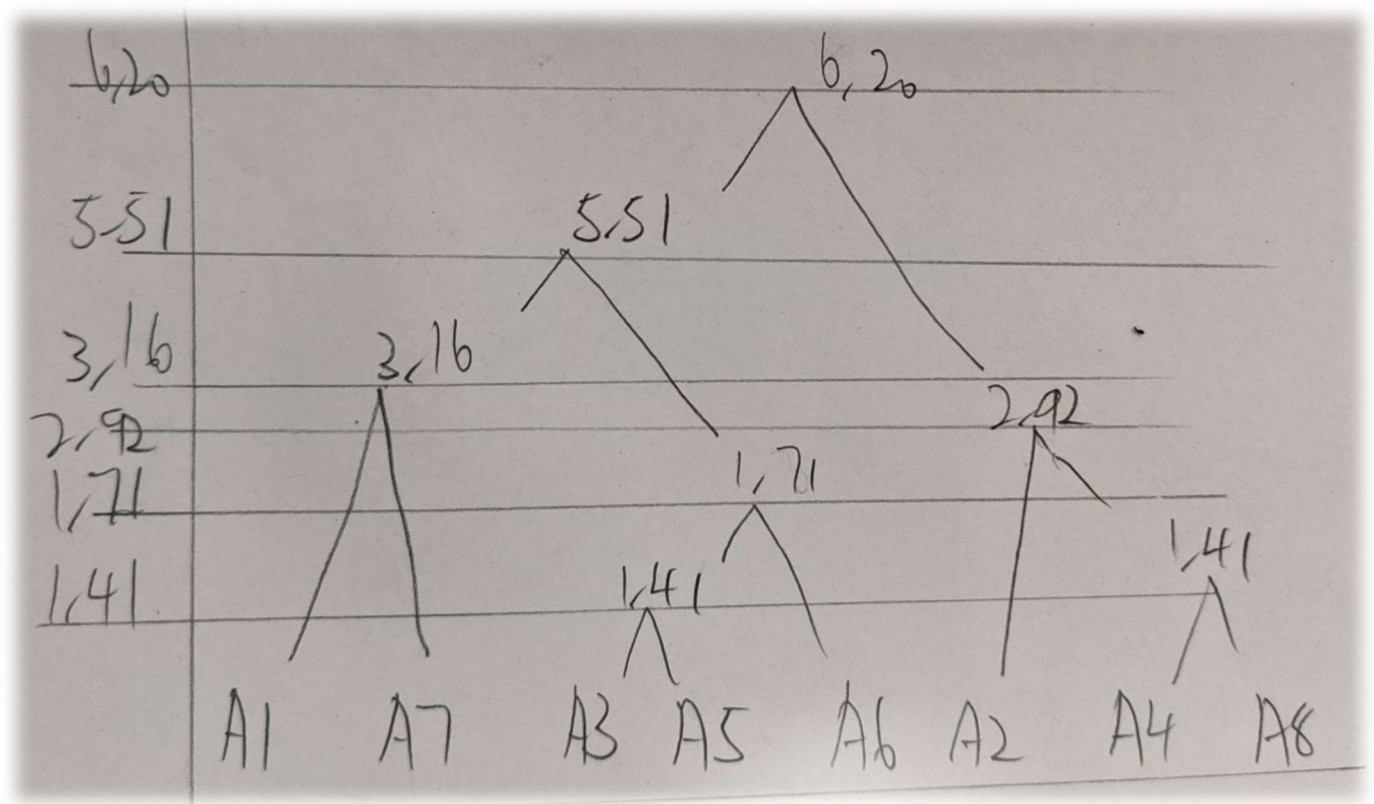
	A1	(A2, (A4, A8))	((A3, A5), A6)	A7
A1	0	4.678695	4.832244	3.162278
(A2, (A4, A8))		0	6.186396	7.737848
((A3, A5), A6)			0	6.189661
A7				0

	(A1, A7)	(A2, (A4, A8))	((A3, A5), A6)
(A1, A7)	0	6.208272	5.510953
(A2, (A4, A8))		0	6.186396
((A3, A5), A6)			0

	((A1, A7), ((A3, A5), A6))	(A2, (A4, A8))
((A1, A7), ((A3, A5), A6))	0	6.197334
(A2, (A4, A8))		0

	((((A1, A7), ((A3, A5), A6)), (A2, (A4, A8))))
((((A1, A7), ((A3, A5), A6)), (A2, (A4, A8))))	0

Dendrograms



Question 3a

∴ SVM is trying to maximize the margin between 2 classes

∴ The optimal decision boundary is original and cross the point $b(2.5, 3.5)$

It perpendicular to the line between support vectors $(2, 3)$ and $(3, 4)$

$$\text{slope: } (-1) \frac{4-3}{3-2} = -1$$

$$\text{line equation: } (x_2 - 3.5) = -1(x_1 - 2.5)$$

$$x_1 + x_2 = 6$$

Weight Vector:

$$2w_1 + 3w_2 + b = 1$$

$$3w_1 + 4w_2 + b = -1$$

Question 3b

