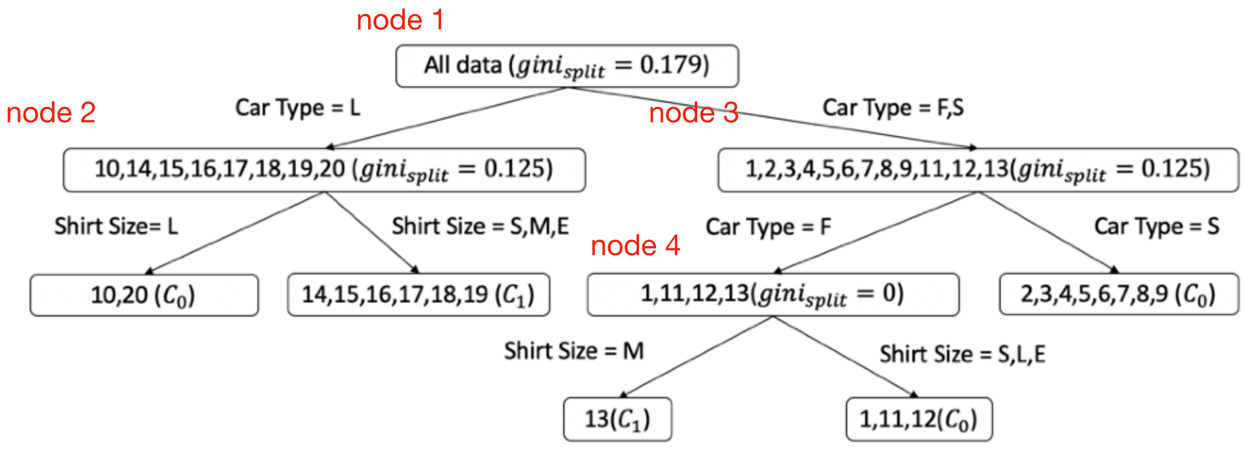
HW5 r08921a07 曾梓豪
Problem 1:
tree:



All split and its corresponding gini:
node 1:
Split with Gender (['F'], ['M']), gini: 0.3999999999999999
Split with Car Type (['Family'], ['Sport', 'Luxury']), gini: 0.46875
Split with Car Type (['Sport'], ['Family', 'Luxury']), gini: 0.2666666666666666
Split with Car Type (['Luxury'], ['Family', 'Sport']), gini: 0.17916666666666672
Split with Shirt Size (['Samll'], ['Medium', 'Large', 'Extra Large']), gini: 0.48
Split with Shirt Size (['Medium'], ['Samll', 'Large', 'Extra Large']), gini: 0.44835164835164837
Split with Shirt Size (['Large'], ['Samll', 'Medium', 'Extra Large']), gini: 0.46875
Split with Shirt Size (['Extra Large'], ['Samll', 'Medium', 'Large']), gini: 0.46875
Split with Shirt Size (['Samll', 'Medium'], ['Large', 'Extra Large']), gini: 0.45
Split with Shirt Size (['Samll', 'Large'], ['Medium', 'Extra Large']), gini: 0.47272727272727283
Split with Shirt Size (['Samll', 'Extra Large'], ['Medium', 'Large']), gini: 0.47272727272727283
node 2:
Split with Gender (['F'], ['M']), gini: 0.21428571428571433
Split with Shirt Size (['Samll'], ['Medium', 'Large', 'Extra Large']), gini: 0.20833333333333326
Split with Shirt Size (['Medium'], ['Samll', 'Large', 'Extra Large']), gini: 0.19999999999999999
Split with Shirt Size (['Large'], ['Samll', 'Medium', 'Extra Large']), gini: 0.125
Split with Shirt Size (['Extra Large'], ['Samll', 'Medium', 'Large']), gini: 0.21428571428571433
Split with Shirt Size (['Samll', 'Medium'], ['Large', 'Extra Large']), gini: 0.16666666666666666
Split with Shirt Size (['Samll', 'Large'], ['Medium', 'Extra Large']), gini: 0.1875
Split with Shirt Size (['Samll', 'Extra Large'], ['Medium', 'Large']), gini: 0.19999999999999999
node 3:
Split with Gender (['F'], ['M']), gini: 0.14814814814814817
Split with Car Type (['Family'], ['Sport', 'Luxury']), gini: 0.125
Split with Car Type (['Sport'], ['Family', 'Luxury']), gini: 0.125
Split with Shirt Size (['Samll'], ['Medium', 'Large', 'Extra Large']), gini: 0.14814814814814817
Split with Shirt Size (['Medium'], ['Samll', 'Large', 'Extra Large']), gini: 0.125
Split with Shirt Size (['Large'], ['Samll', 'Medium', 'Extra Large']), gini: 0.14999999999999994
Split with Shirt Size (['Extra Large'], ['Samll', 'Medium', 'Large']), gini: 0.14814814814814817
Split with Shirt Size (['Samll', 'Medium'], ['Large', 'Extra Large']), gini: 0.1428571428571429
Split with Shirt Size (['Samll', 'Large'], ['Medium', 'Extra Large']), gini: 0.1428571428571429
Split with Shirt Size (['Samll', 'Extra Large'], ['Medium', 'Large']), gini: 0.13888888888888884
node 4:
Split with Shirt Size (['Samll'], ['Medium', 'Large', 'Extra Large']), gini: 0.33333333333333337
Split with Shirt Size (['Medium'], ['Samll', 'Large', 'Extra Large']), gini: 0.0
Split with Shirt Size (['Large'], ['Samll', 'Medium', 'Extra Large']), gini: 0.33333333333333337
Split with Shirt Size (['Extra Large'], ['Samll', 'Medium', 'Large']), gini: 0.33333333333333337
Split with Shirt Size (['Samll', 'Medium'], ['Large', 'Extra Large']), gini: 0.25
Split with Shirt Size (['Samll', 'Large'], ['Medium', 'Extra Large']), gini: 0.25
Split with Shirt Size (['Samll', 'Extra Large'], ['Medium', 'Large']), gini: 0.25

Problem 2:

$$P(c=0 \mid G=F, C=L, S=L) = \frac{P(G=F, C=L, S=L \mid c=0)P(c=0)}{P(G=F, G=L, S=L)}$$

$$= \frac{P(G=F \mid c=0)P(C=L \mid c=0)P(S=L \mid c=0)P(c=0)}{P(G=F, C=L, S=L)}$$

$$= \frac{\frac{4}{12}\frac{1}{12}\frac{3}{12}\frac{6}{10}}{\frac{1}{10}} = \frac{72}{12^3}$$

$$P(c=1 \mid G=F, C=L, S=L) = \frac{P(G=F, C=L, S=L \mid c=0)P(c=1)}{P(G=F, G=L, S=L)}$$

$$= \frac{P(G=F \mid c=1)P(C=L \mid c=1)P(S=L \mid c=1)P(c=1)}{P(G=F, C=L, S=L)}$$

$$= \frac{\frac{6}{8}\frac{7}{8}\frac{1}{8}\frac{4}{10}}{\frac{1}{10}} = \frac{168}{8^3}$$

→ the prediction of Naive Bayes classifier is 1

Problem 3:

minimuze objective $\dfrac{|w|^2}{2}$

subject to:

$$(w^T \cdot [4,3]^T + b) - 1 > 0 \quad (w^T \cdot [2,1]^T + b) + 1 < 0 \quad (w^T \cdot [-1,-2]^T + b) + 1 < 0$$
$$(w^T \cdot [7,2]^T + b) - 1 > 0 \quad (w^T \cdot [2,-1]^T + b) + 1 < 0$$
$$(w^T \cdot [4,8]^T + b) - 1 > 0 \quad (w^T \cdot [-1,3]^T + b) + 1 < 0$$

$\rightarrow$

$$([b,w]^T \cdot [-1,-4,-3]^T) < -1 \quad ([b,w]^T \cdot [1,2,1]^T) < -1$$
$$([b,w]^T \cdot [-1,-7,-2]^T) < -1 \quad ([b,w]^T \cdot [1,2,-1]^T) < -1$$
$$([b,w]^T \cdot [-1,-4,-8]^T) < -1 \quad ([b,w]^T \cdot [1,-1,3]^T) < -1$$
$$([b,w]^T \cdot [1,-1,-2]^T) < -1$$

Let $P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $q = 0$ and $h = -1$

Use python library “qpsolvers” to solve the quadratic programming problem below.

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T P x + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b \\ & lb \leq x \leq ub \end{array}$$

we can get the hyperplane:

$$w = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$
$$b = -2.5$$
$$f(x) = w^T x + b = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}^T \cdot x - 2.5$$

by calculating $f(x)$

$$f(\begin{bmatrix} 4 \\ 3 \end{bmatrix}) = 1,\ f(\begin{bmatrix} 7 \\ 2 \end{bmatrix}) = 2, f(\begin{bmatrix} 4 \\ 8 \end{bmatrix}) = 3.5, f(\begin{bmatrix} 2 \\ 1 \end{bmatrix}) = -1, f(\begin{bmatrix} 2 \\ -1 \end{bmatrix}) = -2, f(\begin{bmatrix} -1 \\ 3 \end{bmatrix}) = -1, f(\begin{bmatrix} -1 \\ -2 \end{bmatrix}) = -4$$

we know two support vectors are:

$$\begin{bmatrix} 4 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$