دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف دوم درس مبانی یادگیری ماشین (بخش تئوری)

نام و نام خانوادگی: علیرضا ابره فروش

شماره دانشجویی: ۹۸۱۶۶۰۳

نیم‌سال تحصیلی: پاییز ۱۴۰۱

مدرّس: دکتر مهران صفایانی

۱

۱.۱ الف

$$\theta = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

مرز تصمیم $\theta^t x = 0$ است. پس داریم:

$$\begin{aligned} &\theta^t x = 0 \\ &\Rightarrow (3) \times (1) + (0) \times (x_1) + (-1) \times (x_2) = 0 \\ &\Rightarrow x_2 = 3 \end{aligned}$$

مرز تصمیم $x_2 = 3$ است.

۲.۱ ب

$$\theta = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

مرز تصمیم $\theta^t x = 0$ است. پس داریم:

$$\begin{aligned} &\theta^t x = 0 \\ &\Rightarrow (-2) \times (1) + (1) \times (x_1) + (1) \times (x_2) = 0 \\ &\Rightarrow x_2 = -x_1 + 2 \end{aligned}$$

مرز تصمیم $x_2 = -x_1 + 2$ است.

۳.۱ ج

$$\begin{aligned} &z \geq 0 : \quad 0.5 \leq \sigma(z) \leq 1 \\ &z \leq 0 : \quad 0 \leq \sigma(z) \leq 0.5 \\ &\Rightarrow \theta^t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \leq 0 \\ &0 \leq \theta^t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\ &0 \leq \theta^t \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \end{aligned}$$

$$0 \leq \theta^t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\Rightarrow \theta_0 + \theta_1 \times 0 + \theta_2 \times 0 \leq 0$$

$$\theta_0 + \theta_1 \times 0 + \theta_2 \times 1 \geq 0$$

$$\theta_0 + \theta_1 \times 1 + \theta_2 \times 0 \geq 0$$

$$\theta_0 + \theta_1 \times 1 + \theta_2 \times 1 \geq 0$$

$$\Rightarrow \theta = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

۲

$$z = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{K-1} \end{bmatrix}, \sigma\left(z\right) = \begin{bmatrix} \frac{e^{z_0}}{\sum_{i=1}^{K-1} e^{z_i}} \\ \frac{e^{z_1}}{\sum_{i=1}^{K-1} e^{z_i}} \\ \vdots \\ \frac{e^{z_{K-1}}}{\sum_{i=1}^{K-1} e^{z_i}} \end{bmatrix}$$

$$\forall i \in \{0, \cdots, K-1\}$$

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=0}^{K-1} e^{z_j}}$$

$$J = \begin{bmatrix} \frac{\partial \sigma(z)_0}{\partial z_0} & \cdots & \frac{\partial \sigma(z)_0}{\partial z_{K-1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma(z)_{K-1}}{\partial z_0} & \cdots & \frac{\partial \sigma(z)_{K-1}}{\partial z_{K-1}} \end{bmatrix}$$

$$\forall i, j \in \{0, \cdots, K-1\}$$

$$\frac{\partial}{\partial z_j} \ln\left(\sigma(z)_i\right) = \frac{1}{\sigma(z)_i} \cdot \frac{\partial \sigma(z)_i}{\partial z_j}$$

$$\Rightarrow J_{ij} = \frac{\partial \sigma(z)_i}{\partial z_j} = \sigma(z)_i \frac{\partial}{\partial z_j} \ln\left(\sigma(z)_i\right) = \sigma(z)_i \frac{\partial}{\partial z_j} \left(z_i - \ln\left(\sum_{k=0}^{K-1} e^{z_k}\right)\right)$$

$$= \sigma(z)_i \left[1\left\{i=j\right\} - \frac{1}{\sum_{k=0}^{K-1} e^{z_k}} \cdot e^{z_j}\right] = \sigma(z)_i \left[1\left\{i=j\right\} - \sigma(z)_j\right]$$

$$\Rightarrow J = \begin{bmatrix} \sigma(z)_0\left(1-\sigma(z)_0\right) & -\sigma(z)_0\sigma(z)_1 & \cdots & -\sigma(z)_0\sigma(z)_{K-1} \\ -\sigma(z)_1\sigma(z)_0 & \sigma(z)_1\left(1-\sigma(z)_1\right) & \cdots & -\sigma(z)_1\sigma(z)_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma(z)_{K-1}\sigma(z)_0 & -\sigma(z)_{K-1}\sigma(z)_1 & \cdots & \sigma(z)_{K-1}\left(1-\sigma(z)_{K-1}\right) \end{bmatrix} \mathcal{L}\left(y\right) = -\sum_{i=0}^{K-1} y_i \cdot \ln\left(\sigma\left(z\right)_i\right)$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial z_j} = -\frac{\partial}{\partial z_j} \sum_{i=0}^{K-1} y_i \cdot \ln\left(\sigma\left(z\right)_i\right)$$

$$= -\sum_{i=0}^{K-1} y_i \cdot \frac{\partial}{\partial z_j} \ln\left(\sigma\left(z\right)_i\right)$$

$$= -\sum_{i=0}^{K-1} \frac{y_i}{\sigma(z)_i} \cdot \frac{\partial \sigma(z)_i}{\partial z_j}$$

$$= -\sum_{i=0}^{K-1} \frac{y_i}{\sigma(z)_i} \cdot \sigma\left(z\right)_i \cdot \left(1\left\{i=j\right\} - \sigma\left(z\right)_j\right)$$

$$= -\sum_{i=0}^{K-1} y_i \cdot \left(1\left\{i=j\right\} - \sigma\left(z\right)_j\right)$$

$$= \sum_{i=0}^{K-1} y_i \cdot \sigma\left(z\right)_j - \sum_{i=1}^{K-1} y_i \cdot 1\left\{i=j\right\}$$

$= \sum_{i=0}^{K-1} y_i \cdot \sigma(z)_j - y_j$
$= \sigma(z)_j \cdot \sum_{i=0}^{K-1} y_i - y_j$
$= \sigma(z)_j - y_j$

$\Rightarrow \frac{\partial \mathcal{L}}{\partial z} = \sigma(z) - y$

۳

۱.۳ الف

$x^1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, x^2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, x^3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, y^1 = -1, y^2 = -1, y^3 = 1$
$\mathcal{L}(w, b, \alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^3 \alpha_i \left(y^i \left(x^{i^T} w + b \right) - 1 \right)$
$= \frac{1}{2} w^T w - \sum_{i=1}^3 \alpha_i y^i x^{i^T} w - b \left(\sum_{i=1}^3 \alpha_i y^i \right) + \sum_{i=1}^3 \alpha_i$

$\Rightarrow \frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^3 \alpha_i y^i x^i = 0 \Rightarrow w^* = \sum_{i=1}^3 \alpha_i y^i x^i$
$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^3 \alpha_i y^i = 0$

$\Rightarrow w^* = -\alpha_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \alpha_2 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \alpha_3 \begin{bmatrix} 4 \\ 5 \end{bmatrix}$
$-\alpha_1 - \alpha_2 + \alpha_3 = 0$

$\Rightarrow \max_{\alpha_1,\alpha_2,\alpha_3} \mathcal{L}(w, b, \alpha) = \max_{\alpha_1,\alpha_2,\alpha_3} \left[\sum_{i=1}^3 \alpha_i - \frac{1}{2} \sum_{i,j=1}^3 y^i y^j \alpha_i \alpha_j x^{i^T} x^j - b \left(\sum_{i=1}^3 \alpha_i y^i \right) \right]$
$= \max_{\alpha_1,\alpha_2,\alpha_3} \left[\sum_{i=1}^3 \alpha_i - \frac{1}{2} \sum_{i,j=1}^3 y^i y^j \alpha_i \alpha_j x^{i^T} x^j \right] = \max_{\alpha_1,\alpha_2,\alpha_3} \left[-\frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 y^i y^j \alpha_i \alpha_j x^{i^T} x^j + \sum_{i=1}^3 \alpha_i \right]$

$x^{1^T} x^1 = 17, x^{1^T} x^2 = 14, x^{1^T} x^3 = 24, x^{2^T} x^2 = 13, x^{2^T} x^3 = 23, x^{3^T} x^3 = 41$

$\Rightarrow \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2} \left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1\alpha_3 + 13\alpha_2^2 - 46\alpha_2\alpha_3 + 41\alpha_3^2 \right] + \alpha_1 + \alpha_2 + \alpha_3$
$= \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2} \left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1^2 - 48\alpha_1\alpha_2 + 13\alpha_2^2 - 46\alpha_1\alpha_2 - 46\alpha_2^2 + 41\alpha_1^2 + 41\alpha_2^2 + 82\alpha_1\alpha_2 \right] + 2\alpha_1 + 2\alpha_2$
$= \max_{\alpha_1,\alpha_2,\alpha_3} -5\alpha_1^2 - 8x^1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, x^2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, x^3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, y^1 = -1, y^2 = -1, y^3 = 1$
$\mathcal{L}(w, b, \alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^3 \alpha_i \left(y^i \left(x^{i^T} w + b \right) - 1 \right)$

$= \frac{1}{2}w^Tw - \sum_{i=1}^{3}\alpha_i y^i {x^i}^T w - b\left(\sum_{i=1}^{3}\alpha_i y^i\right) + \sum_{i=1}^{3}\alpha_i$

$\Rightarrow \frac{\partial\mathcal{L}}{\partial w} = w - \sum_{i=1}^{3}\alpha_i y^i x^i = 0 \Rightarrow w^* = \sum_{i=1}^{3}\alpha_i y^i x^i$
$\frac{\partial\mathcal{L}}{\partial b} = \sum_{i=1}^{3}\alpha_i y^i = 0$

$\Rightarrow w^* = -\alpha_1\begin{bmatrix}1\\4\end{bmatrix} - \alpha_2\begin{bmatrix}2\\3\end{bmatrix} + \alpha_3\begin{bmatrix}4\\5\end{bmatrix}$
$-\alpha_1 - \alpha_2 + \alpha_3 = 0$

$\Rightarrow \max_{\alpha_1,\alpha_2,\alpha_3}\mathcal{L}(w,b,\alpha) = \max_{\alpha_1,\alpha_2,\alpha_3}\left[\sum_{i=1}^{3}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{3} y^i y^j \alpha_i\alpha_j {x^i}^T x^j - b\left(\sum_{i=1}^{3}\alpha_i y^i\right)\right]$
$= \max_{\alpha_1,\alpha_2,\alpha_3}\left[\sum_{i=1}^{3}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{3} y^i y^j \alpha_i\alpha_j {x^i}^T x^j\right] = \max_{\alpha_1,\alpha_2,\alpha_3}\left[-\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} y^i y^j \alpha_i\alpha_j {x^i}^T x^j + \sum_{i=1}^{3}\alpha_i\right]$

${x^1}^T x^1 = 17, {x^1}^T x^2 = 14, {x^1}^T x^3 = 24, {x^2}^T x^2 = 13, {x^2}^T x^3 = 23, {x^3}^T x^3 = 41$

$\Rightarrow \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2}\left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1\alpha_3 + 13\alpha_2^2 - 46\alpha_2\alpha_3 + 41\alpha_3^2\right] + \alpha_1 + \alpha_2 + \alpha_3$
$= \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2}\left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1^2 - 48\alpha_1\alpha_2 + 13\alpha_2^2 - 46\alpha_1\alpha_2 - 46\alpha_2^2 + 41\alpha_1^2 + 41\alpha_2^2 + 82\alpha_1\alpha_2\right] + 2\alpha_1 + 2\alpha_2$
$= \max_{\alpha_1,\alpha_2,\alpha_3} -5\alpha_1 + 2\alpha_1 + 2\alpha_1 x^1 = \begin{bmatrix}1\\4\end{bmatrix}, x^2 = \begin{bmatrix}2\\3\end{bmatrix}, x^3 = \begin{bmatrix}4\\5\end{bmatrix}, y^1 = -1, y^2 = -1, y^3 = 1$
$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{3}\alpha_i\left(y^i\left({x^i}^T w + b\right) - 1\right)$
$= \frac{1}{2}w^Tw - \sum_{i=1}^{3}\alpha_i y^i {x^i}^T w - b\left(\sum_{i=1}^{3}\alpha_i y^i\right) + \sum_{i=1}^{3}\alpha_i$

$\Rightarrow \frac{\partial\mathcal{L}}{\partial w} = w - \sum_{i=1}^{3}\alpha_i y^i x^i = 0 \Rightarrow w^* = \sum_{i=1}^{3}\alpha_i y^i x^i$
$\frac{\partial\mathcal{L}}{\partial b} = \sum_{i=1}^{3}\alpha_i y^i = 0$

$\Rightarrow w^* = -\alpha_1\begin{bmatrix}1\\4\end{bmatrix} - \alpha_2\begin{bmatrix}2\\3\end{bmatrix} + \alpha_3\begin{bmatrix}4\\5\end{bmatrix}$
$-\alpha_1 - \alpha_2 + \alpha_3 = 0$

$\Rightarrow \max_{\alpha_1,\alpha_2,\alpha_3}\mathcal{L}(w,b,\alpha) = \max_{\alpha_1,\alpha_2,\alpha_3}\left[\sum_{i=1}^{3}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{3} y^i y^j \alpha_i\alpha_j {x^i}^T x^j - b\left(\sum_{i=1}^{3}\alpha_i y^i\right)\right]$

$= \max_{\alpha_1,\alpha_2,\alpha_3} \left[\sum_{i=1}^{3} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{3} y^i y^j \alpha_i \alpha_j x^{i^T} x^j\right] = \max_{\alpha_1,\alpha_2,\alpha_3} \left[-\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} y^i y^j \alpha_i \alpha_j x^{i^T} x^j + \sum_{i=1}^{3}\alpha_i\right]$

$x^{1^T} x^1 = 17, x^{1^T} x^2 = 14, x^{1^T} x^3 = 24, x^{2^T} x^2 = 13, x^{2^T} x^3 = 23, x^{3^T} x^3 = 41$

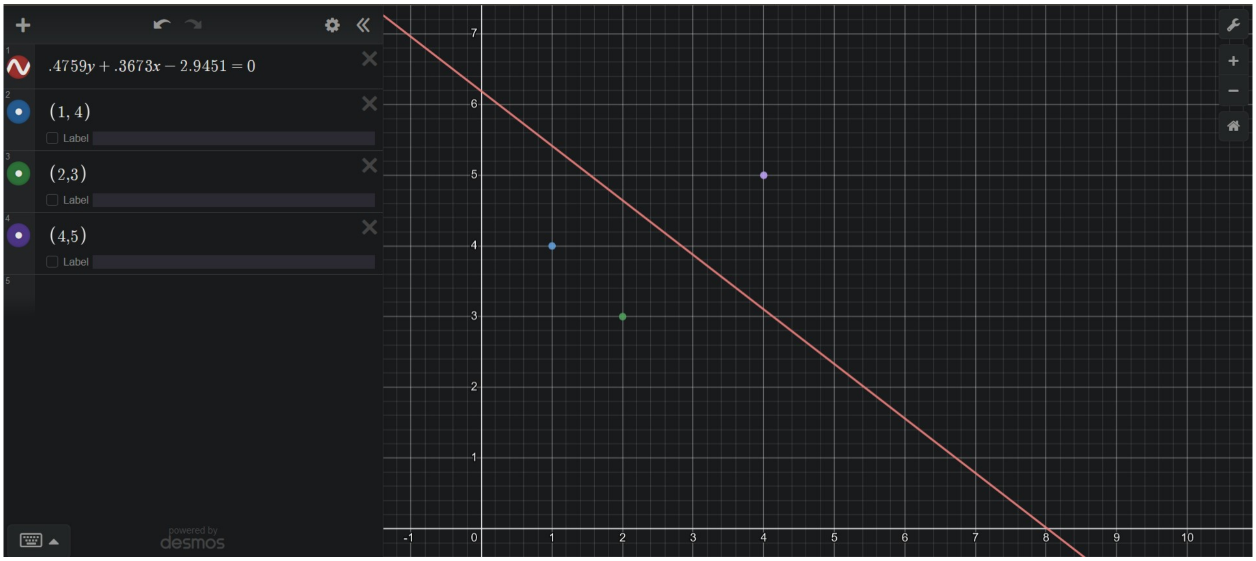
$\Rightarrow \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2}\left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1\alpha_3 + 13\alpha_2^2 - 46\alpha_2\alpha_3 + 41\alpha_3^2\right] + \alpha_1 + \alpha_2 + \alpha_3$
$= \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2}\left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1^2 - 48\alpha_1\alpha_2 + 13\alpha_2^2 - 46\alpha_1\alpha_2 - 46\alpha_2^2 + 41\alpha_1^2 + 41\alpha_2^2 + 82\alpha_1\alpha_2\right] + 2\alpha_1 + 2\alpha_2$
$= \max_{\alpha_1,\alpha_2,\alpha_3} \left[-5\alpha_1^2 - 8\alpha_1\alpha_2 - 4\alpha_2^2 + 2\alpha_1 + 2\alpha_2\right]$

$\Rightarrow J = -5\alpha_1^2 - 8\alpha_1\alpha_2 - 4\alpha_2^2 + 2\alpha_1 + 2\alpha_2$
$\frac{\partial J}{\alpha_1} = -10\alpha_1 - 8\alpha_2 + 2 = 0 \Rightarrow \alpha_1 = -0.8\alpha_2 + 0.2$
$\frac{\partial J}{\alpha_2} = -8\alpha_1 - 8\alpha_2 + 2 = 0 \Rightarrow 0.64\alpha_2 - 1.6 - 8\alpha_2 + 2 = 0$
$\Rightarrow \alpha_2 \simeq 0.0543, \alpha_1 \simeq 0.1565, \alpha_3 = 0.2108$

$$w^* \simeq -0.0543 \begin{bmatrix}1\\4\end{bmatrix} - 0.1565 \begin{bmatrix}2\\3\end{bmatrix} + 0.2108 \begin{bmatrix}4\\5\end{bmatrix} = \begin{bmatrix}0.4759\\0.3673\end{bmatrix}$$
$$b^* = -\begin{bmatrix}1\\4\end{bmatrix}^T \begin{bmatrix}0.4759\\0.3673\end{bmatrix} - b = 1 \Rightarrow b = -2.9451$$

$\Rightarrow 0.4759x_2 + 0.3673x_1 - 2.9451 = 0$

۲.۳ ب



شکل ۱

۴

۱.۴ الف

$$G(D) = 1 - \sum_{i=1}^{k} p_i^2 = 1 - \left(p_{c0}^2 + p_{c1}^2\right)^2 = 1 - \left[\left(\frac{10}{20}\right)^2 + \left(\frac{10}{20}\right)^2\right] = \frac{1}{2}$$

۲.۴ ب

$$\forall i \in \{1, 2, \cdots, 20\}: \quad G_{CustomerID}(D_i) = 1 - \sum_{j=1}^{k} p_j^2 = 1 - \left(\frac{1}{1}\right)^2 = 0$$

$$G_{CustomerID}(D) = 0$$

۳.۴ ج

$$G_{Gender}(D_M) = 1 - \left[\left(\frac{6}{10}\right)^2 + \left(\frac{4}{10}\right)^2\right] = 0.48$$

$$G_{Gender}(D_F) = 1 - \left[\left(\frac{4}{10}\right)^2 + \left(\frac{6}{10}\right)^2\right] = 0.48$$

$$\frac{|D_M|}{D} = \frac{10}{20} = \frac{1}{2}$$

$$\frac{|D_F|}{D} = \frac{10}{20} = \frac{1}{2}$$

$$G_{Gender}(D) = \frac{1}{2} \times 0.48 + \frac{1}{2} \times 0.48 = 0.48$$

۴.۴ د

$$G_{CarType}(D_{Family}) = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right] = 0.375$$

$$G_{CarType}(D_{Sport}) = 1 - \left[\left(\frac{8}{8}\right)^2\right] = 0$$

$G_{CarType}\left(D_{Luxury}\right) = 1 - \left[\left(\frac{1}{8}\right)^2 + \left(\frac{7}{8}\right)^2\right] = 0.21875$
$\frac{|D_{Family}|}{D} = \frac{4}{20} = 0.2$
$\frac{|D_{Sport}|}{D} = \frac{8}{20} = 0.4$
$\frac{|D_{Luxury}|}{D} = \frac{8}{20} = 0.4$
$G_{CarType}\left(D\right) = 0.2 \times 0.375 + 0.4 \times 0 + 0.4 \times 0.21875 = 0.1625$

۵.۴ و

$G_{ShirtSize}\left(D_{Small}\right) = 1 - \left[\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right] = 0.48$
$G_{ShirtSize}\left(D_{Medium}\right) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = \frac{24}{49}$
$G_{ShirtSize}\left(D_{Large}\right) = 1 - \left[\left(\frac{3}{7}\right)^2 + \left(\frac{4}{7}\right)^2\right] = 0.5$
$G_{ShirtSize}\left(D_{ExtraLarge}\right) = 1 - \left[\left(\frac{3}{7}\right)^2 + \left(\frac{4}{7}\right)^2\right] = 0.5$
$\frac{|D_{Small}|}{D} = \frac{5}{20} = 0.25$
$\frac{|D_{Medium}|}{D} = \frac{7}{20} = 0.35$
$\frac{|D_{Large}|}{D} = \frac{4}{20} = 0.2$
$\frac{|D_{ExtraLarge}|}{D} = \frac{4}{20} = 0.2$
$G_{ShirtSize}\left(D\right) = 0.25 \times 0.48 + 0.35 \times \frac{24}{49} + 0.2 \times 0.5 + 0.2 \times 0.5 \simeq 0.4915$

۶.۴ ه

$G_{CarType}\left(D\right) \leq G_{Gender}\left(D\right) \leq G_{ShirtSize}\left(D\right)$

از آنجایی که ویژگیِ Car Type کمترین شاخص Gini را دارد پس مناسب‌تر است.

۷.۴ ی

درست است که ویژگیِ Customer ID دارای کمترین شاخص Gini است. اما چون به ازای هر داده مقدار متمایزی دارد نمی‌توان از آن برای دسته‌بندی استفاده کرد.

۵

۱.۵ الف

$E = -\sum_{i=1}^{k} p_i \log_2 p_i$
$\Rightarrow E = -\left(\frac{4}{9}\log_2\frac{4}{9} + \frac{5}{9}\log_2\frac{5}{9}\right) \simeq -\left[\frac{4}{9}\left(2 - 3.17\right) + \frac{5}{9}\left(2.32 - 3.17\right)\right] \simeq 0.99$

۲.۵ ب

$E\left(D_T\right) = -\left[\frac{3}{4}\ log_2\frac{3}{4} + \frac{1}{4}\ log_2\frac{1}{4}\right] \simeq 0.815$
$E\left(D_F\right) = -\left[\frac{1}{5}\ log_2\frac{1}{5} + \frac{4}{5}\ log_2\frac{4}{5}\right] \simeq 0.72$
$E_{a_1}\left(D\right) = \frac{4}{9}E\left(D_T\right) + \frac{5}{9}E\left(D_F\right) \simeq 0.762$

$\alpha\left(a_1, D\right) = E\left(D\right) - E_{a_1}\left(D\right) \simeq 0.99 - 0.762 = 0.228$

$E\left(D_T\right) = -\left[\frac{2}{5}\, log_2 \frac{2}{5} + \frac{3}{5}\, log_2 \frac{3}{5}\right] \simeq 0.972$
$E\left(D_F\right) = -\left[\frac{2}{4}\, log_2 \frac{2}{4} + \frac{2}{4}\, log_2 \frac{2}{4}\right] \simeq 1$
$E_{a_2}\left(D\right) = \frac{5}{9} E\left(D_T\right) + \frac{4}{9} E\left(D_F\right) \simeq 0.984$
$\alpha\left(a_2, D\right) = E\left(D\right) - E_{a_2}\left(D\right) \simeq 0.99 - 0.984 = 0.006$

۳.۵ ج

$E\left(D_1\right) = -\left[\frac{1}{1}\log_2 \frac{1}{1} + \frac{0}{1}\log_2 \frac{0}{1}\right] = 0$
$E\left(D_3\right) = -\left[\frac{0}{1}\log_2 \frac{0}{1} + \frac{1}{1}\log_2 \frac{1}{1}\right] = 0$
$E\left(D_4\right) = -\left[\frac{1}{1}\log_2 \frac{1}{1} + \frac{0}{1}\log_2 \frac{0}{1}\right] = 0$
$E\left(D_5\right) = -\left[\frac{2}{2}\log_2 \frac{2}{2} + \frac{0}{2}\log_2 \frac{0}{2}\right] = 0$
$E\left(D_6\right) = -\left[\frac{1}{1}\log_2 \frac{1}{1} + \frac{0}{1}\log_2 \frac{0}{1}\right] = 0$
$E\left(D_7\right) = -\left[\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{2}\log_2 \frac{1}{2}\right] = 1$
$E\left(D_8\right) = -\left[\frac{0}{1}\log_2 \frac{0}{1} + \frac{1}{1}\log_2 \frac{1}{1}\right] = 0$

$E_{a_3}\left(D_1\right) = \frac{1}{9} E\left(D_1\right) = 0$
$E_{a_3}\left(D_3\right) = \frac{1}{9} E\left(D_3\right) = 0$
$E_{a_3}\left(D_4\right) = \frac{1}{9} E\left(D_4\right) = 0$
$E_{a_3}\left(D_5\right) = \frac{2}{9} E\left(D_5\right) = 0$
$E_{a_3}\left(D_6\right) = \frac{1}{9} E\left(D_6\right) = 0$
$E_{a_3}\left(D_7\right) = \frac{2}{9} E\left(D_7\right) = \frac{2}{9}$
$E_{a_3}\left(D_8\right) = \frac{1}{9} E\left(D_8\right) = 0$

$\alpha\left(a_3, D_1\right) \simeq 0.99 - 0 = 0.99$
$\alpha\left(a_3, D_3\right) \simeq 0.99 - 0 = 0.99$
$\alpha\left(a_3, D_4\right) \simeq 0.99 - 0 = 0.99$
$\alpha\left(a_3, D_5\right) \simeq 0.99 - 0 = 0.99$
$\alpha\left(a_3, D_6\right) \simeq 0.99 - 0 = 0.99$
$\alpha\left(a_3, D_7\right) \simeq 0.99 - \frac{2}{9} \simeq 0.77$
$\alpha\left(a_3, D_8\right) \simeq 0.99 - 0 = 0.99$

$\alpha\left(a_3, D\right) \simeq 0.99 - \frac{2}{9} \simeq 0.77$

۴.۵ د

با توجه به اینکه بین a_1 و a_2 ویژگیِ a_1 بیشترین information gain را دارد. پس تقسیم‌بندی را بهتر انجام می‌دهد.
با توجه به اینکه بین a_1 و a_3 ویژگیِ a_3 بیشترین information gain را دارد. پس تقسیم‌بندی را بهتر انجام می‌دهد.

۵.۵ و

$\text{classification error rate}_{a_1} = \frac{2}{9}$
$\text{classification error rate}_{a_2} = \frac{5}{9}$

$\text{classification error rate}_{a_1} \leq \text{classification error rate}_{a_2}$

با توجه به اینکه ویژگیِ a_1 میزان خطای طبقه‌بندی کمتری دارد تقسیم‌بندی را بهتر انجام می‌دهد.

۶.۵ ه

$G_{a_1}(D_T) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = \frac{3}{8} = 0.375$
$G_{a_1}(D_F) = 1 - \left[\left(\frac{1}{5}\right)^2 + \left(\frac{4}{5}\right)^2\right] = \frac{8}{25} = 0.32$

$G_{a_1}(D) = \frac{4}{9} \times \frac{3}{8} + \frac{5}{9} \times \frac{8}{25} \simeq 0.35$

$G_{a_2}(D_T) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right] = \frac{12}{25} = 0.48$
$G_{a_2}(D_F) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = \frac{1}{2} = 0.5$

$G_{a_2}(D) = \frac{5}{9} \times \frac{12}{25} + \frac{4}{9} \times \frac{1}{2} \simeq 0.49$

با توجه به اینکه بین a_1 و a_2 ویژگیِ a_1 کمترین شاخص Gini را دارد. پس تقسیم‌بندی را بهتر انجام می‌دهد.

۶

به طور شهودی با هر بخش‌بندی، داده‌ها همگن‌تر می‌شوند. در نتیجه آنتروپی افزایش نمی‌یابد. در ادامه به اثبات این قضیه می‌پردازیم.

$E(D) = -\sum_{i=1}^{k} p_i \log_2 p_i$
$\Rightarrow E_{x_i}(y) = -\sum_{j=1}^{k} p(y_j|x_i) \log_2 p(y_j|x_i)$
$E_x(y) = -\sum_{i=1}^{n_x} p(x_i) E_{x_i}(y) = -\sum_{i=1}^{n_x}\sum_{j=1}^{k} p(x_i) p(y_j|x_i) \log_2 p(y_j|x_i) = -\sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 p(y_j|x_i)$
$\Rightarrow E_x(y) - E(y) = -\sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 p(y_j|x_i) - \left[-\sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 p(y_j)\right]$
$= \sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 \frac{p(y_j)}{p(y_j|x_i)} = \sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 \frac{p(y_j)}{\frac{p(y_j,x_i)}{p(x_i)}} = \sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 \frac{p(y_j)p(x_i)}{p(y_j,x_i)}$

$\leq \log_2\left[\sum_{i=1}^{n_x}\sum_{j=1}^{k} p(y_j, x_i) \log_2 \frac{p(y_j)p(x_i)}{p(y_j,x_i)}\right] = \log_2\left[\sum_{i=1}^{n_x} p(x_i) \sum_{j=1}^{k} p(y_j)\right] = \log_2 1 = 0$

$\Rightarrow E_x\left(y\right) \leq E\left(y\right)$

۷

$\min\limits_{\beta_0,\beta} \sum_{i=1}^{N} \left[1 - y_i f\left(x_i\right)\right]_+ + \frac{\lambda}{2} \left\|\beta\right\|^2 \overset{\lambda=\frac{1}{C}}{=} \min\limits_{\beta_0,\beta} \sum_{i=1}^{N} \left[1 - y_i f\left(x_i\right)\right]_+ + \frac{1}{2C} \left\|\beta\right\|^2$
$\underset{\times C}{\Rightarrow} \min\limits_{\beta_0,\beta} \frac{1}{2} \left\|\beta\right\|^2 + C \sum_{i=1}^{N} \left[1 - y_i f\left(x_i\right)\right]_+ \overset{1-y_i f(x_i)\leq\zeta_i}{=} \min\limits_{\beta_0,\beta} \frac{1}{2} \left\|\beta\right\|^2 + C \sum_{i=1}^{N} \left[\zeta_i\right]_+$
$\overset{\zeta_i\geq 0}{=} \min\limits_{\beta_0,\beta} \frac{1}{2} \left\|\beta\right\|^2 + C \sum_{i=1}^{N} \zeta_i$

منابع