

$$\text{Entropy} = - \sum_{i=1}^c -P_i \log_2 P_i$$

سوال ①
(الف)

$$\begin{aligned} H(y) &= -P(w) \log_2 P(w) - P(L) \log_2 P(L) \\ &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1 \end{aligned}$$

$$\text{Gain}(S, A) = H_S(Y) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} H_{S_v}(Y) \quad (\text{ب})$$

$$\begin{aligned} \text{Gain}(S, x_1) &= 1 - \left[\frac{6}{10} \times \left(-\frac{4}{6} \log \frac{4}{6} - \frac{2}{6} \log \frac{2}{6} \right) \right] \\ &\quad - \left[\frac{4}{10} \times \left(-\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} \right) \right] = 0.124 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, x_2) &= 1 - \left[\frac{4}{10} \times \left(-\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} \right) \right] \\ &\quad - \left[\frac{6}{10} \times \left(-\frac{4}{6} \log \frac{4}{6} - \frac{2}{6} \log \frac{2}{6} \right) \right] = 0.124 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, x_3) &= 1 - \left[\frac{5}{10} \times \left(-\frac{4}{5} \log \frac{4}{5} - \frac{1}{5} \log \frac{1}{5} \right) \right] \\ &\quad - \left[\frac{5}{10} \times \left(-\frac{4}{5} \log \frac{4}{5} - \frac{1}{5} \log \frac{1}{5} \right) \right] = 0.278 \end{aligned}$$

چون Gain برای فیچر x_3 بیشتر از بقیه شد پس بهترین است با فیچر x_3 Split انجام شود.

$$\begin{aligned}
 \text{Cost function} = \text{MSE} &= \frac{1}{N} \sum_{i=1}^N (p_i - l_i)^2 \\
 &= \frac{1}{N} \sum_{i=1}^N (m_T(x_i) - y_i)^2 \\
 &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{t=1}^T \beta_t G_t(x_i) - y_i \right)^2
 \end{aligned}$$

ب) با توجه به عبارت تابع هزینه بالا، آن را بصورت باز شده $G_t(x_i)$ می نویسیم

$$J = \text{Cost} = \frac{1}{N} \sum_{i=1}^N \left(\underbrace{\sum_{t=1}^{T-1} \beta_t G_t(x_i)}_{\text{حالت گذرایان نسبت به } \beta_T \text{ می نویسیم}} + \beta_T G_T(x_i) - y_i \right)^2$$

$$\frac{\partial J}{\partial \beta_t} = \frac{1}{N} \sum_{i=1}^N 2 (m_{T-1} + \beta_T G_T(x_i) - y_i) G_T(x_i)$$

$$\Rightarrow \frac{\partial J}{\partial \beta_t} = 0 \Rightarrow \sum_{i=1}^N m_{T-1}(x_i) G_T(x_i) + \beta_t \sum_{i=1}^N G_T(x_i)^2 = \sum_{i=1}^N y_i G_T(x_i)$$

$$\Rightarrow \beta_T = \frac{\sum_{i=1}^N y_i G_T(x_i) - \sum_{i=1}^N m_{T-1}(x_i) G_T(x_i)}{\sum_{i=1}^N G_T(x_i)^2}$$

الف) احتمال انتخاب نشدن داده α در N' بار نمونه گیری برابر است با :

$$\left(\frac{N-1}{N}\right)^N = \left(\frac{N-1}{N}\right)^{PN}$$

اگر $N \rightarrow \infty$ باشد آنگاه :

$$\left(\frac{N-1}{N}\right)^{PN} = e^{-P}$$

چون N داده داریم پس $\underline{e^{-P} \cdot N}$ تا داده انتخاب نخواهد شد.

ب) حالت هایی که random forest حداقل یک جواب درست دهد را حساب می کنیم :

$$(1-0.15)(1-0.25)(1-0.35) = 0.414 \quad (1) \text{ همه درست}$$

$$(1-0.15)(1-0.25)0.35 + (1-0.15)0.25(1-0.35) + 0.15(1-0.25)(1-0.35) \quad (2) \text{ دو تا درست}$$

$$(1-0.15)0.25 \times 0.35 + 0.15 \times (1-0.25) \times 0.35 + 0.15 \times 0.25 \times (1-0.35) \quad (3) \text{ یکی درست}$$

(4) E_{out} خواهد شد :

$$E_{out}(G) = 1 - \text{جمع سه حالت بالا}$$

سوال 4

کترینه الف و د درست است و بقیه غلط می باشد.

آموزش در مدل bagging بصورت مجزا توسط مدل های مختلف بصورت موازی روی مجموعه دیتاست انتخابی انجام می گردد.

در boosting به دنبال بهینه کردن نتایج مدل مرحله قبل روی داده ها هستیم پس نسبت به مدل های مختلف وابسته است امکان مولی اجرا شدن ندارند همچنین نتایج روی مدل داده آموزش بهینه می شود.

سوال 5

الف) به ازای $K=1$ ، خطای آموزش مدل بهینه و صفر خواهد بود.
اما این نشان دهنده overfit است و نشان دهنده عملکرد خوب روی داده test نیست.

ب) مقادیر کوچک K ، باعث اور فیت شدن مدل روی داده های آموزش می شود.
مقادیر بزرگ K ، باعث underfit شدن مدل شدن و بایس زیادی دارد.
پس overfit بدلیل واریانس بالا و underfit بدلیل بایس بالا مناسب نیستند.

ج) روش LOOCV

به ازای $K=1$ $\left\{ \begin{array}{l} 5 \text{ غلط برای کلاس منفی} \\ 5 \text{ غلط برای کلاس مثبت} \end{array} \right\}$ خطا: $\frac{10}{14}$

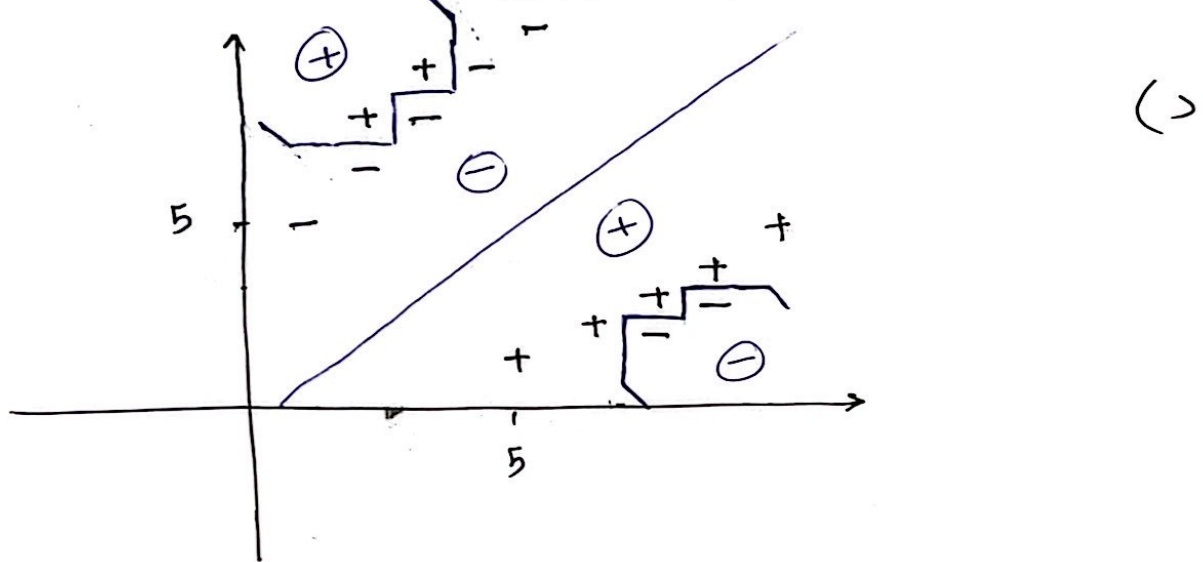
به ازای $K=3$ $\left\{ \begin{array}{l} 3 \text{ غلط برای کلاس منفی} \\ 3 \text{ غلط برای کلاس مثبت} \end{array} \right\}$ خطا: $\frac{6}{14}$

سوال 5

ادامه ج) به ازای $K=5$ ← $\left. \begin{array}{l} 2 \text{ خطی برای کلاس منفی} \\ 2 \text{ خطی برای کلاس مثبت} \end{array} \right\}$ خطا: $\frac{4}{14}$

به ازای $K=7$ و بالاتر هم همین مقدار خطی وجود دارد و کمینه ترین حالت است.

پس به ازای $K=5$ or 7 or ... برقرار است.



سوال 6

$$\left\{ \begin{array}{l} H(T) = - \sum_{i=1}^K P_{C_i} \log P_{C_i} \\ H(T) = - \sum_{i=1}^K \frac{1}{K} \log \frac{1}{K} = - \log_2 \frac{1}{K} = \log_2 K \end{array} \right.$$

$$\left\{ \begin{array}{l} H(T|A) = \sum_{i=1}^{m_A} H(T_i) P(A_i) = \sum_{i=1}^{m_A} \log K \times \frac{1}{m_A} \\ \Rightarrow H(T|A) = m_A \times \log K \times \frac{1}{m_A} = \log_2 K \end{array} \right.$$

$$IG(T, A) = H(T) - H(T|A) = \log_2 K - \log_2 K = 0$$