به نام خدا

علی قاسم زاده ۴۰۱۱۰۶۳۳۹

SUBJECT:

Year: Month: Day:

Page : ( )

۱- ابتدا IG مربوط به x_1, x_2, x_3 مربوط به درس machine learning را حساب می‌کنیم و بین آنها max می‌گیریم:

برای تبدیل شدن به دو دسته، جاهای مساوی یا بزرگتر مساوی را بررسی می‌کنیم

$$IG(x_1) = H(Y) - H(Y|x_1)$$
$$IG(x_2) = H(Y) - H(Y|x_2)$$
$$IG(x_3) = H(Y) - H(Y|x_3)$$

$$H(Y) = -\sum_{i=1}^{k} P(Y=y_i)\log_2 P(Y=y_i) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} \approx 0.97$$

$$H(Y|X_1) = -\sum_{j=1}^{v} P(X_1 \geq \overset{6,7,8}{\uparrow}{x_j}) \sum_{i=1}^{k} P(Y=y_i|X_1 \geq x_j)\log_2 P(Y=y_i|X_1\geq x_j) =$$

برای ۶: $-\frac{1}{5}\log 1 - \frac{4}{5}\times\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = -\frac{4}{5}\log_2\frac{1}{2} = \frac{4}{5}\log_2 2 = \frac{4}{5} = 0.8 \rightarrow IG(Y,X_1)_6 = 0.17$

برای ۷: $-\frac{2}{5}(1\times\log 1) - \frac{3}{5}\times\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right) = 0.55 \rightarrow IG(Y,X_1)_7 = 0.42$

برای ۸: $-\frac{3}{5}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) - \frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 0.95 \rightarrow IG(Y,X_1)_8 = 0.02$

$$H(Y|X_2) = -\sum_{j=1}^{v} P(X_2 = x_j)\sum_{i=1}^{k} P(Y=y_i|X_2=x_j)\log_2 P(Y=y_i|X_2=x_j) = \ldots$$

$$-\frac{3}{5}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) - \frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 0.95 \rightarrow IG(Y,X_2) = 0.02$$

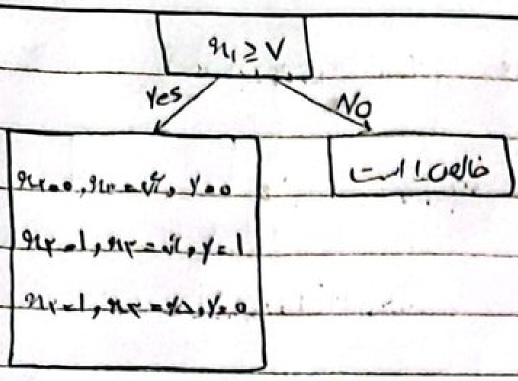
$$H(Y|X_3) = -\sum_{j=1}^{v} P(X_3 \geq x_j)\sum_{i=1}^{k} P(Y=y_i|X_3\geq x_j)\log_2 P(Y=y_i|X_3\geq x_j)$$

برای ۰/۵: $-\frac{1}{5}(1\times\log 1) - \frac{4}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 0.8 \rightarrow IG(Y,X_3)_{0.5} = 0.17$

برای ۰/۶: $-\frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) - \frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 0.95 \rightarrow IG(Y,X_3)_{0.6} = 0.02$

$IG(Y,X_1)_7$ بیشترین مقدار است پس دسته‌بندی براساس آن صورت می‌گیرد

Senobar



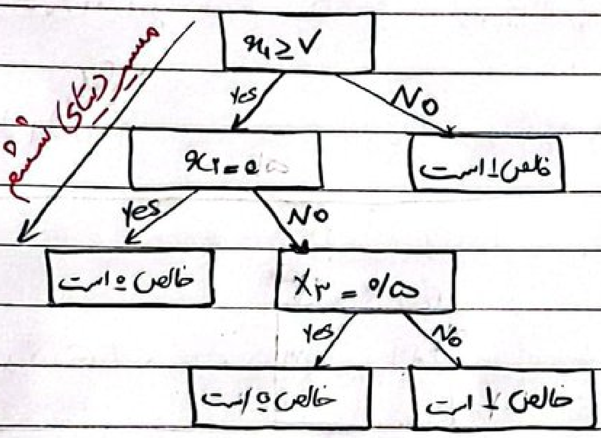
$$H(Y) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0/92$$

$$H(Y|X_2) = -\frac{1}{3}(\log_2 1) - \frac{2}{3}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = \frac{2}{3} \rightarrow IG = 0/25$$

طبق فرض سوال X_2 را انتخاب می‌کنیم

$$H(Y|X_3) = -\frac{1}{3}(\log_2 1) - \frac{2}{3}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = \frac{2}{3} \quad IG = 0/25$$

به دو حالت داریم اینم با اینز حساب می‌شود



منظور Y خالص صفر است و Y خالص یک است.

تمدید نمی‌شود

حالا برای درس معماری کامپیوتر حساب می‌کنیم.

$$H(Y) = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5} = 0/97$$

$$H(Y|X_1) = -\sum_{j=1}^{7} P(X_1 \ge x_j) \sum_{i=1}^{k} P(Y=y_i|X_1 \ge x_j)\log_2 P(Y=y_i|X_1 \ge x_j)$$

برای ۶، $-\frac{1}{5}\log 1 - \frac{4}{5}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 0/8 \longrightarrow IG(Y,X_1)_6 = 0/17$

برای ۷، $-\frac{2}{5}\log_2 1 - \frac{3}{5}\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right) = 0/55 \longrightarrow IG(Y,X_1)_7 = 0/42$

برای ۸، $-\frac{4}{5}\log\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) - \frac{1}{5}\log(1) = 0/8 \longrightarrow IG(Y,X_1)_8 = 0/17$

$$H(Y|X_2) = -\sum_{j=1}^{v} P(X_2=x_j) \sum_{i=1}^{k} P(Y=y_i|X_2=x_j) \log_2 P(Y=y_i|X_2=x_j)$$

$$= -\frac{2}{5}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) - \frac{3}{5}\left(\frac{1}{3}\log_2\frac{1}{3}+\frac{2}{3}\log_2\frac{2}{3}\right) = 0.95 \longrightarrow IG(Y,X_2) = 0.02$$

$$H(Y|X_3) = -\sum_{j=1}^{v} P(X_3 \geq x_j) \sum_{i=1}^{k} P(Y=y_i|X_3 \geq x_j) \log_2 P(Y=y_i|X_3 \geq x_j)$$

برای ۰٫۳ : $-\frac{2}{5}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) - \frac{3}{5}\left(\frac{1}{3}\log_2\frac{1}{3}+\frac{2}{3}\log_2\frac{2}{3}\right) = 0.95 \longrightarrow IG(Y,X_3)_{0.3} = 0.02$

برای ۰٫۴ : $-\frac{3}{5}\left(\frac{1}{3}\log_2\frac{1}{3}+\frac{2}{3}\log_2\frac{2}{3}\right) - \frac{2}{5}(1\times\log_2 1) = 0.55 \longrightarrow IG(Y,X_3)_{0.4} = 0.42$

برای ۰٫۷ : $-\frac{4}{5}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) - \frac{1}{5}(1\times\log_2 1) = 0.8 \longrightarrow IG(Y,X_3)_{0.7} = 0.17$

بیشترین مقدار مربوط به $IG(Y,X_1)_7$ ، همچنین $IG(Y,X_3)_{0.4}$ است که طبق فرض سوال

اولی انتخاب می‌شود.

$X_1 \geq 7$

yes:
$x_2=1, x_3=0.3, Y=1$
$x_2=1, x_3=0.5, Y=0$
$x_2=0, x_3=0.5, Y=1$

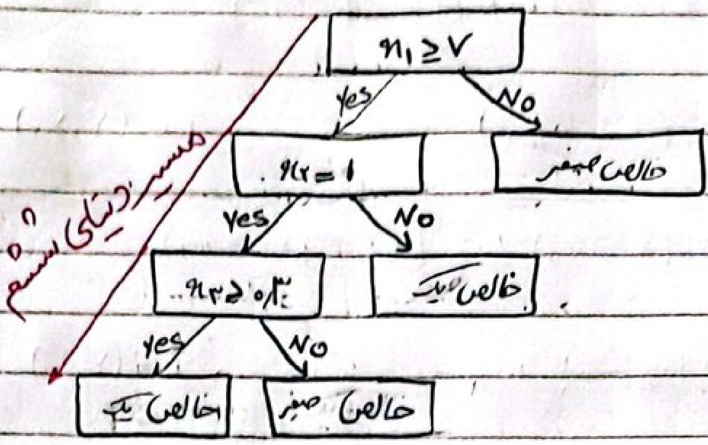
no: خالص صفر

$$H(Y) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.92$$

$$H(Y|X_2) = -\frac{1}{3}\times\log 1 - \frac{2}{3}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) = \frac{2}{3} \simeq 0.67 \longrightarrow IG(Y,X_2) = 0.25$$

$$H(Y|X_3) = -\frac{1}{3}\times\log 1 - \frac{2}{3}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) = \frac{2}{3} \simeq 0.67 \longrightarrow IG(Y,X_3) = 0.25$$

طبق فرض سوال $IG(Y,X_2)$ انتخاب می‌شود



→ تمدید می‌شود

← اولویت با کمترین m1 است.

۲-آ)

$$P(Y=1 \mid X_1=1, X_2=0, X_3=0) \propto P(Y=1) \times \prod_{i=1}^{3} P(X_i=x_i \mid Y=1) = 0/5 \times p \times q \times q = \frac{pq^2}{2}$$

$$P(Y=0 \mid X_1=1, X_2=0, X_3=0) \propto P(Y=0) \times \prod_{i=1}^{3} P(X_i=x_i \mid Y=1) = 0/5 \times (1-p)(1-q)(1-q) = \frac{(1-p)(1-q)^2}{2}$$

باید مورد اول بزرگتر از مورد دوم باشد تا $Y=1$ انتخاب شود

$$\frac{pq^2}{2} > \frac{(1-p)(1-q)^2}{2} \rightarrow$$

$$pq^2 > (1-p)(q^2-2q+1) \rightarrow pq^2 > q^2-2q+1-pq^2+2pq-p \rightarrow$$

$$2pq^2-2pq+p > q^2-2q+1 \rightarrow p(2q^2-2q+1) > (1-q)^2 \rightarrow \boxed{p > \frac{(1-q)^2}{2q^2-2q+1}}$$

ب) x_3 هم تکراری x_2 هست یعنی با احتمال برابر است با ...

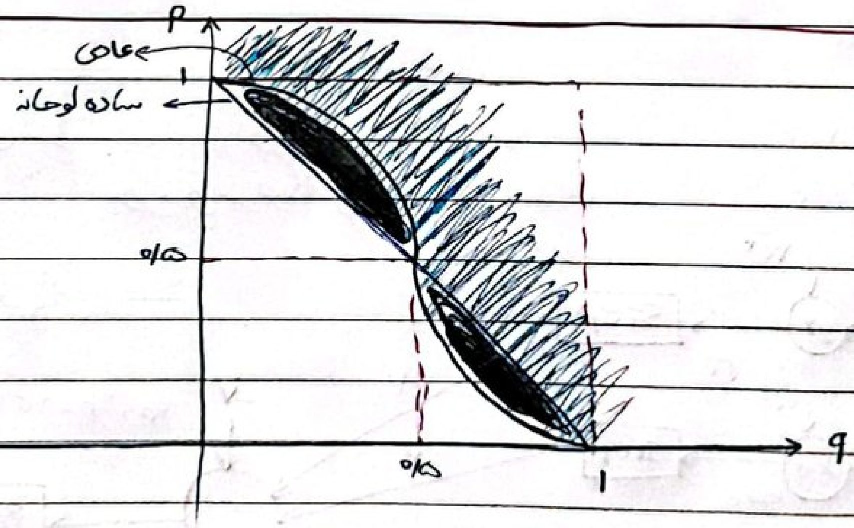
$$P(X_1=1, X_2=X_3=0 \mid Y=1) = P(X_1=1 \mid Y=1)\,P(X_2=0 \mid Y=1) = pq$$

$$P(X_1=1, X_2=X_3=0 \mid Y=0) = P(X_1=1 \mid Y=0)\,P(X_2=0 \mid Y=0) = (1-p)(1-q)$$

باید مورد اول بزرگتر از مورد دوم باشد تا $Y=1$ انتخاب شود

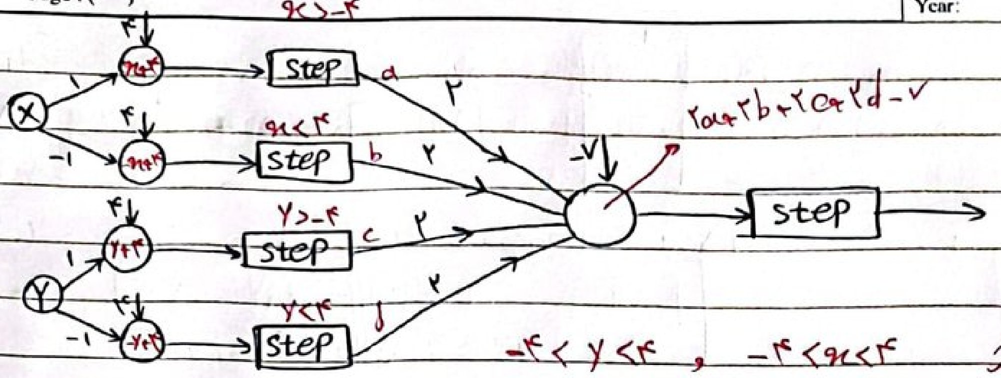
$$pq > (1-p)(1-q) \rightarrow pq > 1-q-p+pq \rightarrow \boxed{p > 1-q}$$

ج)



در دو ناحیه سایه، بیز عادی و بیز ساده لوحانه جواب های متفاوتی دارند ← در این دو ناحیه خطا داریم

۳_ الف)



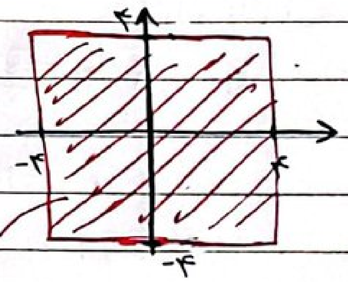
بدین ترتیب اگر $-4<x<4$ و $-4<y<4$

باشد، ۸ وارد دایره می‌شود و ۸−۷=۱ و خروجی تابع step به ازای ۱ برابر است با

۱) و در غیر اینصورت حتی اگر یکی از شرط‌ها نقض شوند ورودی دایره کوچکتر مساوی

۶ خواهد بود و ورودی step منفی خواهد بود ← خروجی step صفر خواهد بود.

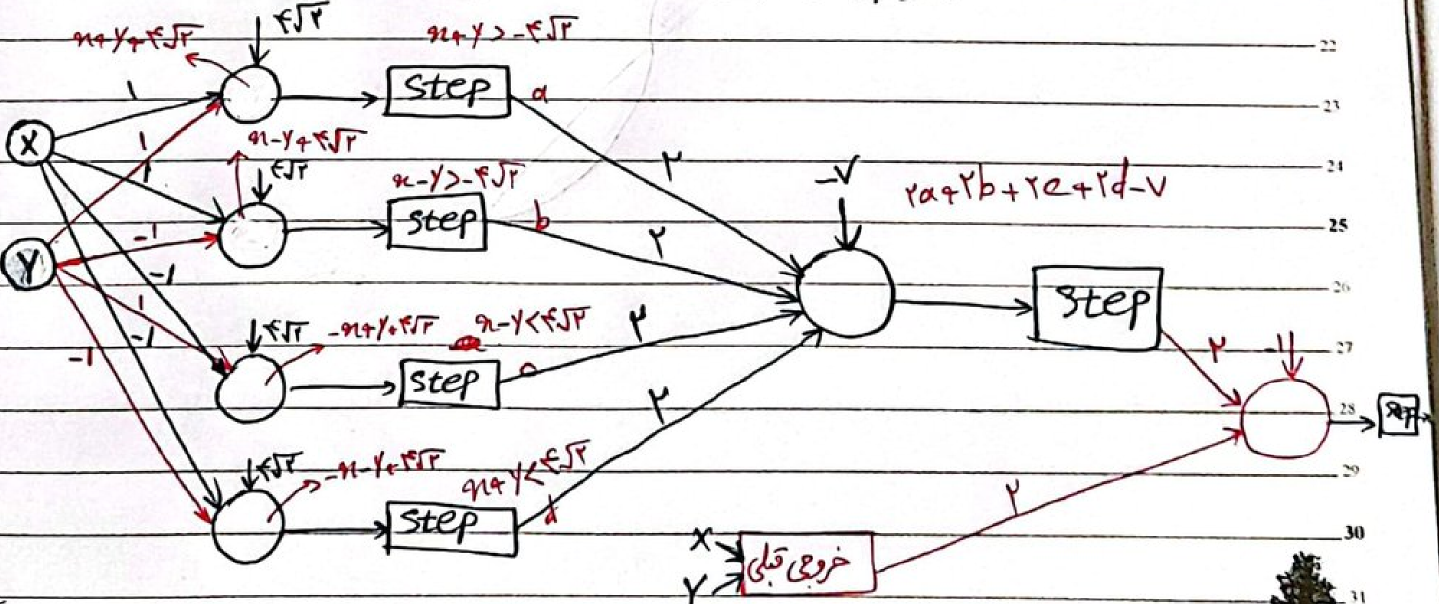
شبکه عصبی → $\begin{cases} 1 & \text{if } -4<x<4 \text{ و } -4<y<4 \\ 0 & \text{else} \end{cases}$



این ناحیه یک، بقیه جاها صفر

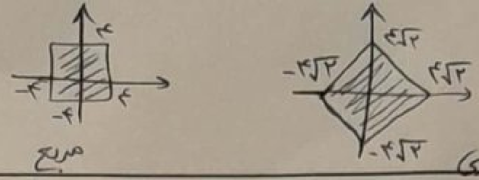
ب) این شرط برای لوزی در شکل استفاده می‌شود $-4\sqrt{2} < x+y < 4\sqrt{2}$

برای مربع هم از بخش الف استفاده می‌کنیم $-4\sqrt{2} < x-y < 4\sqrt{2}$



بدین ترتیب اگر در مربع بالایی باشد یا در لوزی باشد، ۸ در ورودی step بزرگتر از صفر خواهد بود و خروجی step برابر یک خواهد بود و اگر هم

الف)



هر دو تا نقض شده باشند، یعنی نقطه مورد نظر نه در لوزی افتاده و نه در مربع

پس ورودی ~~step~~ دایره آخری ۰x۲+۰x۲ خواهد بود و ۱- می‌شود ←

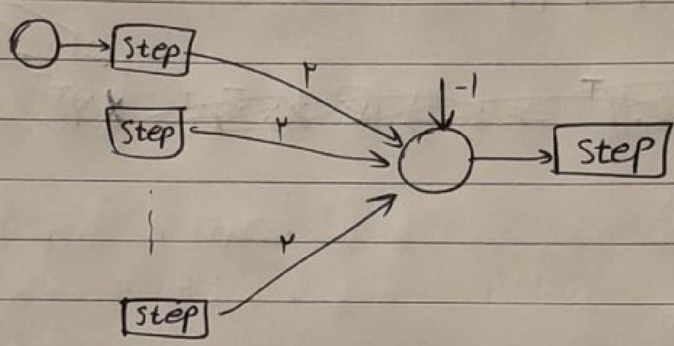
ورودی استپ منفی خواهد بود ← خروجی استپ صفر خواهد بود.

ج) مربع اولیه را تعداد خوبی بار دوران می‌دهیم و برای هر کدام یک شبکه عصبی می‌نویسیم و

در نهایت همه را or می‌کنیم خروجی همه را در ۲ ضرب می‌کنیم و حاصل جمعش را

منهای 1 می‌کنیم، اگر در هر کدام افتاده بود step آخری یک می‌دهد وگرنه صفر می‌دهد.

خروجی‌های شبکه‌های عصبی



مثلاً ۰ درجه دوران، ۱ درجه دوران، ... ، ۹۰ درجه دوران را در نظر می‌گیریم و معادله خط‌هایش

را در می‌آوریم مانند بخش ب معادله‌های تعدادی مربع به دست می‌آید که هر یک را با یک شبکه عصبی

پیاده می‌کنیم و در نهایت از همه مانند شکل بالا اجتماع می‌گیریم و خروجی step آخرمان می‌شود

جواب تقریبی مان.

۴- الف)

$$y_i = \sum_{i=1}^{m} \beta_j x_{ij} + \beta_0 = [\beta_0 \cdots \beta_m] \begin{bmatrix} x_{i1} \\ \vdots \\ x_{im} \end{bmatrix} = \beta^T \cdot x_i$$

$$L = \sum_{i=1}^{n} (\overset{x_i \cdot \beta}{\quad} - y_i)^2 = \left\| \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & \cdots & x_{2m} \\ \vdots & \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \right\|_2 =$$

$$(X\beta - y).(X\beta - y)$$

ب)

$$\frac{\partial L}{\partial \beta} = 2X.(X\beta - y) = 2X^TX\beta - 2X^Ty = 0 \longrightarrow X^TX\beta = X^Ty$$

$$\longrightarrow \beta = (X^TX)^{-1}X^Ty$$

می‌خواهیم L را کمینه کنیم پس مشتق می‌گیریم و برابر صفر قرار می‌دهیم

ج)

$$L = \sum_{i=1}^{n} (\overset{x_i.\beta}{\cancel{\beta x}} - y_i)^2 + \lambda \sum_{i=0}^{m} \beta_i^2 = (X\beta - y)(X\beta - y) + \lambda\beta^T\beta \longrightarrow$$

$$\frac{\partial L}{\partial \beta} = 2X^TX\beta - 2X^Ty + 2\lambda\beta = 0 \longrightarrow X^TX\beta + \overset{\lambda\beta}{\cancel{\beta}} = X^Ty \longrightarrow$$

$$(X^TX + \lambda I)\beta = X^Ty \longrightarrow \boxed{\beta = (X^TX + \lambda I)^{-1}X^Ty}$$

از L نسبت به β مشتق می‌گیریم و برابر صفر قرار می‌دهیم

SUBJECT

Year: Month: Day:

۵- w^* را آن wای می‌گیریم که B را تولید می‌کند

$$\to \langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle = \langle w^*, y_i x_i\rangle = y_i\langle w^*, x_i\rangle \geq 1$$

$$\langle w^*, w^{0}\rangle = \langle w^*, (0, 0, \dots, 0)\rangle = 0$$

$$\langle w^*, w^{(T+1)}\rangle = \langle w^*, w^{(T+1)}\rangle - \langle w^*, w^T\rangle + \langle w^*, w^T\rangle - \langle w^*, w^{(T-1)}\rangle +$$

$$\langle w^*, w^{(T-1)}\rangle - \dots - \langle w^*, w^0\rangle = \sum_{t=1}^{T}\left(\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle\right)$$

$$= \sum_{t=1}^{T} y_i\langle w^*, x_i\rangle \geq \sum_{t=1}^{T} 1 \Rightarrow$$

$$\boxed{\langle w^*, w^{(T+1)}\rangle \geq T}$$

حالا اگر آپدیت نداشته باشیم که تغییری نداریم و اگر داشته باشیم

$$\|w^{t+1}\|^2 = \|w^t + y_i x_i\|^2 = \|w^t\|^2 + \|y_i x_i\|^2 + 2y_i\langle w^t, x_i\rangle$$

چون -1 یا $+1$ باشد توان دوم‌اش یک است — منفی است

$$\|w^{t+1}\|^2 \leq \|w^t\|^2 + \|x_i\|^2$$ طول هر x_i کمتر مساوی R است طبق سوال

$$\Rightarrow \|w^{t+1}\|^2 \leq \|w^t\|^2 + R^2$$

$$\|w^{(1)}\|^2 \leq R^2 \leftarrow \|w^{(1)}\|^2 \leq \underbrace{\|w^{(0)}\|}_{0} + R^2$$: پایه استقرا ←

$$\|w^{(T+1)}\|^2 \leq TR^2$$ فرض استقرا

$$\|w^{(T+2)}\|^2 \leq (T+1)R^2$$ حکم استقرا

طبق فرض استقرا : $\|w^{(T+2)}\|^2 \leq \|w^{(T+1)}\|^2 + R^2 \leq TR^2 + R^2 = (T+1)R^2$

داریم که ← $\|w^{(T+1)}\|^2 \leq TR^2$ پس $\|w^{(T+1)}\| \leq \sqrt{T}R$

حالا می‌دانیم که $T \leq \langle w^*, w^{(T+1)}\rangle \leq \underbrace{\|w^*\|}_{B} \times \|w^{(T+1)}\|$

Senobar

$\Rightarrow T \le \langle w^*, w^{(T+1)} \rangle \le \|w^{(T+1)}\| \times \|w^*\| \le \sqrt{T} R B$

$\Rightarrow T \le \sqrt{T} R B \to \sqrt{T} \le RB \to T \le (RB)^2$

← حکم مسئله ۲ برقرار است و حداکثر تعداد تکرارها $(RB)^2$ است

همچنین پس از متوقف شدن الگوریتم داریم که:

$\forall i \in [m], \ y_i \langle w^t, x_i \rangle > 0$

چرا که اگر برقرار نباشد یعنی یک i ای داریم که $y_i \langle w^t, x_i \rangle$ منفی است

و در نتیجه در دفعه‌ی Tام وارد if می‌شویم و حلقه برای بار $T+1$ام

اجرا می‌شود که با فرض متوقف شدن الگوریتم در تضاد است.

پس زمانی که حلقه دیگر اجرا نشده باشد یعنی وارد else شده‌ایم و در نتیجه

حکم سوال برقرار است.

ب) قرار می‌دهیم $d=m$ و $x_i = e_i$ که e_i برداری است که فقط درایه‌ی iام آن یک است

و بقیه درایه‌هایش صفر است. $\to R = \max_i \|x_i\| = \max_i 1 = 1$

$w^{(1)} = 0 \to y_1 \langle w^{(1)}, e_1 \rangle = 0$

$w^{(2)} = w^{(1)} + y_1 e_1 = 0 + e_1 = e_1$

فرض کنیم $w^{(t)} = \sum_{i=1}^{t-1} e_i$، حالا اثبات می‌کنیم که $w^{(t+1)} = \sum_{i=1}^{t} e_i$

$y_t \langle w^{(t)}, e_t \rangle = y_t \langle \sum_{i=1}^{t-1} e_i, e_t \rangle = 0 \Rightarrow$

$w^{(t+1)} = w^{(t)} + y_t e_t = \sum_{i=1}^{t-1} e_i + e_t = \sum_{i=1}^{t} e_i$

حال برای $t=m$ داریم که

$$w^{(m+1)} = \sum_{i=1}^{m} e_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \longrightarrow \|w^{(m+1)}\| = m$$

همچنین داریم که $\forall$ $y_i \langle w^{(m+1)}, x_i \rangle = 1 > 0$

پس این الگوریتم m بار اجرا می‌شود و به اتمام می‌رسد.

حال $R=1$ و $B=\sqrt{m}$ (چون فقط به ازای $t=m$ داریم که $y_i \langle w^{(t+1)}, x_i \rangle$ و $\forall i \in [m]$

$$\Rightarrow (RB)^2 = m$$

همچنین در قسمت اول داشتیم که $T \leq (RB)^2$ و $T=m$ است و $(RB)^2 = m$

است یعنی حالت تساوی بیان کوچکتر مساوی برقرار است

که این هم جواب است ← حکم قسمت (الف) هم در این بخش برقرار است.