subject: تمرین هوش مصنوعی

date: رادین سامانی

۴۰۱۱۰۶۰۹۴

Machine Learning → $H(Y) = -۲/۵ \lg ۲/۵ - ۳/۵ \lg ۳/۵ = ۰٫۹۷۱$ ۱

■ x_1 : ۵ ; ۶ ; ۷ ; ۸

① ② ③ : possible splits

$H(Y|①) = -۱/۵ \times ۰ - ۴/۵(۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) = ۰٫۸$

$H(Y|②) = -۲/۵ \times ۰ - ۳/۵(۱/۳ \lg ۱/۳ + ۲/۳ \lg ۲/۳) = ۰٫۵۵۱$ ✓

$H(Y|③) = -۳/۵(۱/۳ \lg ۱/۳ + ۲/۳ \lg ۲/۳) - ۲/۵(۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) = ۰٫۹۵۱$

■ x_2 : $H(Y|x_2) = -۳/۵(۱/۳ \lg ۱/۳ + ۲/۳ \lg ۲/۳) - ۲/۵(۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) = ۰٫۹۵۱$

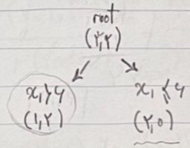
■ x_3 : ۰٫۴ ; ۰٫۵ ; ۰٫۶

① ② : possible splits

$H(Y|①) = -۱/۵ \times ۰ - ۴/۵ \times (۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) = ۰٫۸$

$H(Y|②) = -۳/۵(۱/۳ \lg ۱/۳ + ۲/۳ \lg ۱/۳) - ۲/۵(۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) = ۰٫۹۵۱$

→ pick min of $H(Y|?)$: split on $x_1 > ۶$



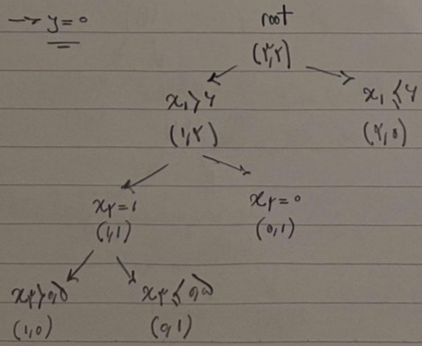
✓ continue on $x_1 > ۶$

■ x_2 : $H(Y|x_2) = -۲/۳(۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) - ۱/۳ \times ۰ = ۲/۳$

■ x_3 : ۰٫۵ ; ۰٫۶ → $H(Y|x_3) = -۲/۳(۱/۲ \lg ۱/۲ + ۱/۲ \lg ۱/۲) = ۲/۳$

→ split on x_2, then split on x_3 : ML Decision Tree :

for $x_1 = ۷, x_2 = ۰, x_3 = ۰٫۵۵$ → $y = ۰$



maxnote

Computer Architecture $\rightarrow H(Y) = ۰٫۹۵۱$

■ x_1 : ۵ ¦ ۶ ¦ ۷ ¦ ۸
① ② ③

$H(Y|①) = -\frac{۱}{۵}\times 0 - \frac{۴}{۵}\times(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) = ۰٫۸$

$H(Y|②) = -\frac{۲}{۵}\times 0 - \frac{۳}{۵}\times(\frac{۲}{۳}\lg\frac{۲}{۳} + \frac{۱}{۳}\lg\frac{۱}{۳}) = ۰٫۵۵۱$

$H(Y|③) = -\frac{۴}{۵}\times(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) - \frac{۱}{۵}\times 0 = ۰٫۸$

■ x_2 : $H(Y|x_2) = -\frac{۳}{۵}(\frac{۱}{۳}\lg\frac{۱}{۳} + \frac{۲}{۳}\lg\frac{۲}{۳}) - \frac{۲}{۵}(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) = ۰٫۹۵۱$

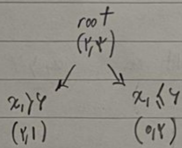
■ x_3 : ۰٫۱ ¦ ۰٫۳ ¦ ۰٫۴ ¦ ۰٫۷
① ② ③

$H(Y|①) = -\frac{۲}{۵}(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) - \frac{۳}{۵}(\frac{۱}{۳}\lg\frac{۱}{۳} + \frac{۲}{۳}\lg\frac{۲}{۳}) = ۰٫۹۵۱$

$H(Y|②) = -\frac{۳}{۵}(\frac{۱}{۳}\lg\frac{۱}{۳} + \frac{۲}{۳}\lg\frac{۲}{۳}) - \frac{۲}{۵}\times 0 = ۰٫۵۵۱$

$H(Y|③) = -\frac{۴}{۵}(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) - \frac{۱}{۵}\times 0 = ۰٫۸$

→ split on $\underline{x_1 \geq ۶}$



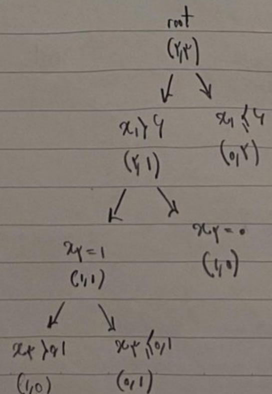
✓ continue on $x_1 \geq ۶$

■ x_2 : $H(Y|x_2) = -\frac{۲}{۳}(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) - \frac{۱}{۳}\times 0 = \frac{۲}{۳}$

■ x_3 : ۰٫۱ ¦ ۰٫۴ $\rightarrow H(Y|x_3) = -\frac{۲}{۳}(\frac{۱}{۲}\lg\frac{۱}{۲} + \frac{۱}{۲}\lg\frac{۱}{۲}) - \frac{۱}{۳}\times 0 = \frac{۲}{۳}$

→ split on x_2, then on x_3 , CA Decision Tree :

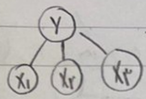
for $x_1 = ۸, x_2 = ۱, x_3 = ۰٫۴$: $\underline{Y = 1}$



subject : date :

۲ (آ

Naive Bayes:



$$P(Y, X_1=1, X_2=0, X_3=0) \rightarrow P(Y=1, \text{features}) = 0.5\,pq^2$$
$$P(Y=0, \text{features}) = 0.5(1-p)(1-q)^2$$

$$\rightarrow P(Y=1 \mid \text{features}) = \frac{P(Y=1, \text{features})}{P(\text{features})} = \frac{pq^2}{pq^2+(1-p)(1-q)^2}$$

$$\text{if } P(Y=1 \mid \text{features}) > 0.5 \rightarrow \underline{Y=1}$$

$$\underline{pq^2 > (1-p)(1-q)^2} \quad \text{Cond } 1$$

ب) بدون مسئله ، احتمال X_3 را در نظر نمی‌گیریم , feature ها X_1 , X_2 است پس حساب کنیم:

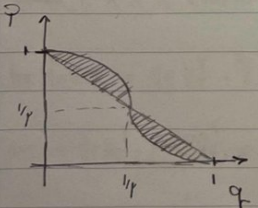
$$P(Y=1, \text{features}) = 0.5\,pq$$
$$P(Y=0, \text{features}) = 0.5(1-p)(1-q)$$
$$P(Y=1 \mid \text{features}) = \frac{P(Y=1, \text{features})}{P(\text{features})} = \frac{pq}{pq+(1-p)(1-q)}$$

$$\text{if } P(Y=1 \mid \text{features}) > 0.5 \rightarrow \underline{Y=1}$$

$$\underline{pq > (1-p)(1-q)} \quad \text{Cond } 2$$

ج)

Error in Naive Bayes: $C_1 \wedge \sim C_2$ or $C_2 \wedge \sim C_1 = C_1 \oplus C_2$



$$\text{Cond } 2 \rightarrow pq = (1-p)(1-q)$$
$$p = 1-q$$

$$\text{Cond } 1 \rightarrow pq^2 = (1-p)(1-q)^2$$

10

منحنی درون شکل

در ناحیه هاشور خورده، Naive Bayes دچار خطا می‌شود!

سوال سوم) آ) ابتدا لایه output را بررسی می‌کنیم. اگر خروجی های لایه hidden را h_1 تا h_4 بنامیم، خروجی این لایه به صورت زیر می‌شود:

$$Output = step(2\sum_{i=1}^{4} h_i - 7)$$

با توجه به اینکه خروجی های h_i، تابع های step هستند، پس برد $\{0,1\}$ دارند و در نتیجه output زمانی فعال می‌شود که تمامی h_iها برابر ۱ بشوند.

حال به بررسی h_i ها می‌پردازیم

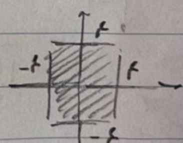
$$h_1 = step(x+4) \rightarrow h_1 = \begin{cases} 1 & x \geq -4 \\ 0 & x < -4 \end{cases}$$

$\vdots$

اگر به همین ترتیب بررسی را ادامه دهیم، برای فعال شدن output، ۴ شرط زیر باید فعال باشد:

1) $x \geq -4$ 2) $-x \geq -4$ 3) $y \geq -4$ 4) $-y \geq -4$

$\longrightarrow |x| \leq 4 \;,\; |y| \leq 4 \longrightarrow$ مربعی به مرکز $(0,0)$ و ضلع های به طول ۸



این شبکه عصبی زمانی فعال می‌شود که مختصات (x,y) درون این مربع هاشورخورده باشه.

ب) برای تعمیم این شبکه عصبی که هر شکل محدبی را پوشش دهد، باید معادله خطوط محدب را بدانیم که ضلع های مربع محدب باشند. سپس همانند شبکه عصبی گفته شده، شروط را اعمال کنیم.

شروط برای مربع مورب :

$x \sin 45 + y \cos 45 \leq 2$

$x \sin 45 + y \cos 45 \geq -2$

$x \sin 135 + y \cos 135 \leq 2$

$x \sin 135 + y \cos 135 \geq -2$

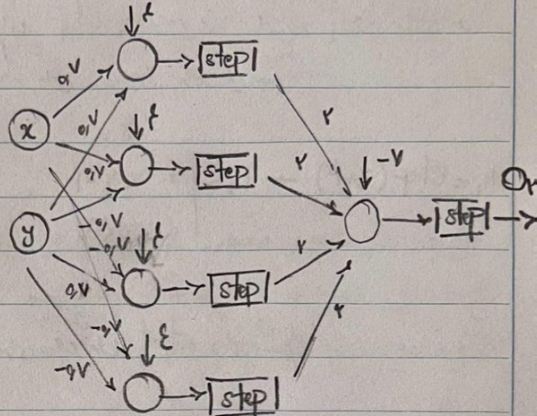
$\downarrow \sqrt{2}/2 = 0.7$

$0.7x + 0.7y \leq 2$

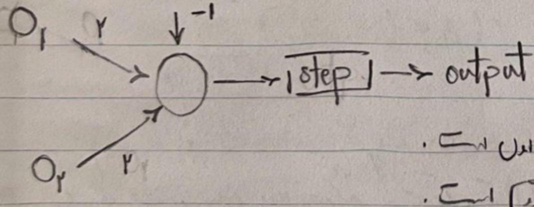
$-0.7x - 0.7y \leq 2$

$-0.7x + 0.7y \leq 2$

$0.7x - 0.7y \leq 2$



در نهایت خروجی این شبکه عصبی را با خروجی شبکه عصبی که بالا رسم کرده است or می‌کنیم.
به این منظور این کار را می‌کنیم:



به طوری که O_1 خروجی شبکه عصبی اول است.
O_2 خروجی شبکه عصبی دوم است.

ج) به منظور شبیه کردن شکل به یک دایره، از n مربع مورب برای بخش های مختلف شبکه عصبی استفاده می‌کنیم. به منظور انجام این کار، معادلات زیر را برای این مربع ها استفاده می‌کنیم.

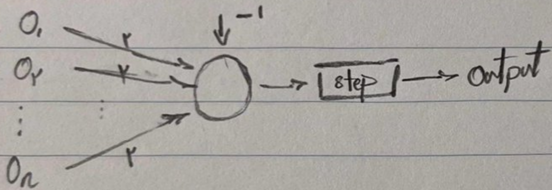
for $i = 1$ to n :

$$|x \sin \theta_i + y \cos \theta_i| \leq \varepsilon$$

$$|x \sin(\theta_i + \pi/2) + y \cos(\theta_i + \pi/2)| = |-x \cos \theta_i + y \sin \theta_i| \leq \varepsilon$$

$$\theta_i = \frac{(i-1)\pi}{2n} \longrightarrow$$ زاویه‌ها از ۰ تا $\frac{\pi}{2}$

سپس در نهایت خروجی هر شبکه عصبی ساخته شده همانند شکل ب را با یکدیگر or می‌کنیم.



subject : date :

erorr

Dataset : $\{y_i, x_{i1}, \dots, x_{im}\}_{i=1}^{n} \longrightarrow y_i = \sum_j \beta_j x_{ij} + \beta_0 + e_i$ (آ ۴

for $i = 1 \sim n$

let $x_i = [1, x_{i1}, \dots, x_{im}]^T$

$\beta = [\beta_0, \beta_1, \dots, \beta_m]^T$

$\longrightarrow y_i = \beta^T x_i + e_i$ for $i = 1 \sim n$

let X be matrix $= [X_1, \dots, X_n]^T = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}$

$Y = [y_1, \dots, y_n]^T$

$e = [e_1, \dots, e_n]^T$

$\Longrightarrow$ equation in matrix form : $Y = X\beta + e$ with loss function $\|e\|_2^2$

خواسته مسئله $\longrightarrow \underset{\beta}{\operatorname{argmin}} \|e\|_2^2 = \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|_2^2$ (ب

$\|Y - X\beta\|_2^2 = (Y - X\beta)^T (Y - X\beta) = (Y^T - \beta^T X^T)(Y - X\beta) = Y^T Y + \beta^T X^T X \beta - 2\beta^T X^T Y$

$\frac{d}{d\beta} \|Y - X\beta\|_2^2 = 2X^T X\beta - 2X^T Y = 0 \longrightarrow$ for argmin β

$\longrightarrow X^T X \hat{\beta} = X^T Y$

lemma : $X^T X$ is inversible $\longrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$

loss function $= \|e\|_2^2 + \lambda \|\beta\|_2^2$ (ج

$\underset{\beta}{\operatorname{argmin}} \|e\|_2^2 + \lambda \|\beta\|_2^2 = \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$

$\|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 = Y^T Y + \beta^T X^T X \beta - 2\beta^T X^T Y + \lambda \beta^T \beta$

$\frac{d}{d\beta}$ loss function $= 2X^T X \beta - 2X^T Y + 2\lambda \beta = 0 \longrightarrow$ for argmin β

$\longrightarrow (X^T X + \lambda I)\hat{\beta} = X^T Y$

$\longrightarrow \hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y$

Sabject :

سوال ۵ (آ)

let w^* be $\underset{w}{\operatorname{argmin}} \{ \|w\| : \forall i \in [m], y_i \langle w, x_i \rangle \geq 1 \}$

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

$= y_i \langle w^*, x_i \rangle \geq 1$ (on the def. of w^*)

if update occurs

$w^{(1)} = (0, \dots, 0) \to \langle w^*, w^{(1)} \rangle = 0$

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

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

منفی است چون update داریم.

$\to \|w^{(t+1)}\|^2 \leq \|w^{(t)}\|^2 + R^2 \to \|w^{(T+1)}\|^2 \leq TR^2$

$\|w^{(T+1)}\| \leq \sqrt{T} R$ ②

③ $\|w^*\| = B$

$\xrightarrow{3,2,1} \dfrac{\langle w^*, w^{(T+1)} \rangle}{\|w^*\| \, \|w^{(T+1)}\|} \geq \dfrac{T}{B \times \sqrt{T} R} = \dfrac{\sqrt{T}}{RB} \to 1 \geq \dfrac{\sqrt{T}}{RB}$

طبق نامساوی کوشی شوارتز، از ۱ کوچکتر است.

$\to (RB)^2 \geq T$

همان حکم سوال ✓

ب) مثال می‌دهیم $d=m$ و دیتاست زیر را ارائه می‌کنیم: $\{(e_i, y_i=1)\}_{i=1}^{m}$

همچنین بدون کاهش عمومیت، مقدار می‌دهیم $Sign(0)=-1$

اگر الگوریتم را شروع کنیم و w_i را بردار وزن‌ها در مرحله i ام الگوریتم در نظر بگیریم، داریم:

$$w_i = \sum_{j<i} e_j$$

زیرا در هر مرحله، مقدار $y_i\langle w_i, x_i\rangle$ مقدار صفر می‌شود و به بردار w_i، مقدار $y_i x_i = e_i$ اضافه می‌شود. در نتیجه بردار w_i که ابتدا دارای m درایه صفر است، به ترتیب در m مرحله پر می‌شود.

در نهایت داریم w پس از m مرحله به بردار تمام ۱ تبدیل می‌شود. این بردار را w_m می‌نامیم.

✓ همگرایی پرسپترون $\longleftarrow$ $\|w_m\|^2 = m \;,\; \forall i,\; y_i\langle w_m, x_i\rangle = 1 > 0$

می‌دانیم طبق بخش آ که تعداد مراحل که با T نشان می‌دهیم از $(RB)^2$ کم است. در این مسئله، داریم:

$\sqrt{m} \geq B$

$1 \geq R$

$(RB)^2 \geq m$

$(آ) \rightarrow (RB)^2 \geq T$

$(RB)^2$ کران مناسبی است.

$T=m$

برای این مسئله خاص: $T = m$