در هر مرحله با استفاده از information gain بهترین ویژگی (ویژگی ای با بیشترین IG) را برای split کردن داده‌ها انتخاب می‌کنیم. برای این کار به این صورت عمل می‌کنیم که آنتروپی مربوط به Y و همچنین $Y|x_i$ ها را به دست می‌آوریم و با تفاضل این دو به IG می‌رسیم:

$$H(Y) = -\sum_{i=0}^{N} P(Y=y_i)\log_2 P(Y=y_i)$$

$$H(Y|X) = -\sum_{j=0}^{M} P(X=x_j)\sum_{i=0}^{N} P(Y=y_i|X=x_j)\log_2 P(Y=y_i|X=x_j)$$

$$IG(X,Y) = H(Y) - H(Y|X)$$

البته می‌توان صرفا $H(Y|X)$ ویژگی‌ها را حساب کرد و آن ویژگی‌ای را که آنتروپی کمتری دارد را برای split کردن استفاده کرد.

در این صورت برای درس معماری خواهیم داشت:

$$H(y|x_1) = \left\{\begin{array}{l} -P(x_1 \geq j)\sum_{i=0}^{1} P(y=i|x_1\geq j)\log_2 P(y=i|x_1\geq j) \\ -P(x_1<j)\sum_{i=0}^{1} P(y=i|x_1<j)\log_2 P(y=i|x_1<j)\end{array}\right. ;\ j\in\{5,6,7\}\Bigg\} = \{0.80, \boxed{0.55}, 0.80\}$$

$$H(y|x_2) = -\sum_{j=0}^{1} P(x_2=j)\sum_{i=0}^{1} P(y=i|x_2=j)\log_2 P(y=i|x_2=j) = 0.95$$

$$H(y|x_3) = \left\{\begin{array}{l} -P(x_3 \geq j)\sum_{i=0}^{1} P(y=i|x_3\geq j)\log_2 P(y=i|x_3\geq j) \\ -P(x_3<j)\sum_{i=0}^{1} P(y=i|x_3<j)\log_2 P(y=i|x_3<j)\end{array}\right. ;\ j\in\{0.1, 0.3, 0.4\}\Bigg\} = \{0.95, 0.55, 0.80\}$$

ویژگی x_1 با threshold برابر با 6 را در ریشه قرار می‌دهیم.

در شاخه‌ی مربوط به $x_1 \leq 6$ به طور کامل داریم $y=0$ (زیرا از هر دو داده‌ی موجود هر دو تای آن‌ها تمدید نشده‌اند).

در شاخه‌ی $x_1 > 6$ مجددا عمل split را انجام می‌دهیم:

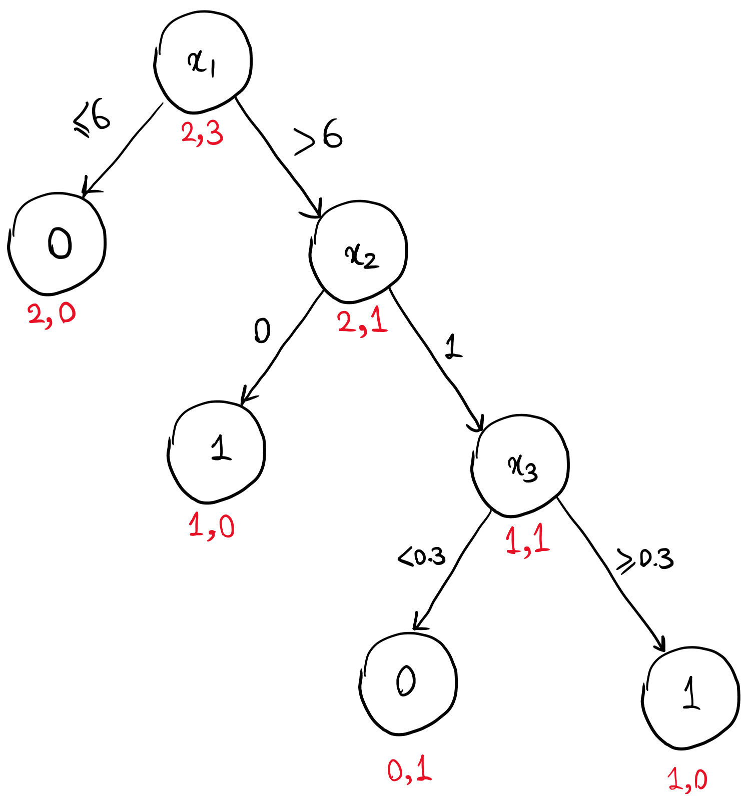
$$H(y|x_2) = -\sum_{j=0}^{1} P(x_2=j)\sum_{i=0}^{1} P(y=i|x_2=j)\log_2 P(y=i|x_2=j) = \boxed{0.67}$$

$$H(y|x_3) = \left\{\begin{array}{l} -P(x_3 \geq j)\sum_{i=0}^{1} P(y=i|x_3\geq j)\log_2 P(y=i|x_3\geq j) \\ -P(x_3<j)\sum_{i=0}^{1} P(y=i|x_3<j)\log_2 P(y=i|x_3<j)\end{array}\right. ;\ j\in\{0.1, 0.3\}\Bigg\} = \{0.67, 0.67\}$$

ویژگی x_2 را برای split کردن استفاده می‌کنیم:

در شاخه‌ای که $x_2=0$ است داریم $y=1$ (چون یک داده داریم و آن تمدید می‌شود).

در شاخه‌ی $x_2=1$ باید دوباره split کردن را انجام دهیم:



با توجه به اطلاعات داده شده و با استفاده از درخت نتیجه می‌گیریم که تمرین درس معماری کامپیوتر احتمالا تمدید می‌شود.

همین کار را برای درس یادگیری ماشین انجام می‌دهیم:

الف) با استفاده از ساختار Naive bayse داریم:

$$Y \to X_1,\quad Y \to X_2,\quad Y \to X_3$$

$$\begin{cases} P(Y=1, X_1=1, X_2=0, X_3=0) = P(Y=1)P(X_1=1\mid Y=1)P(X_2=0\mid Y=1)P(X_3=0\mid Y=1) = \frac{1}{2}pq^2 \\ P(Y=0, X_1=1, X_2=0, X_3=0) = P(Y=0)P(X_1=1\mid Y=0)P(X_2=0\mid Y=0)P(X_3=0\mid Y=0) = \frac{1}{2}(1-p)(1-q)^2 \end{cases}$$

$$\rightsquigarrow P(Y=1\mid X_1=1, X_2=0, X_3=0) = \frac{P(Y=1, X_1=1, X_2=0, X_3=0)}{P(Y=1, X_1=1, X_2=0, X_3=0) + P(Y=0, X_1=1, X_2=0, X_3=0)} = \frac{pq^2}{pq^2 + (1-p)(1-q)^2}$$

ب) از ساختار Naive bayes استفاده نمی‌کنیم:

$$\rightsquigarrow P(Y=1\mid X_1=1, X_2=0, X_3=0) = P(X_1=1, X_2=0, X_3=0\mid Y=1)\frac{P(Y=1)}{P(X_1=1, X_2=0, X_3=0)}$$

$$= P(X_1=1\mid Y=1)P(X_2=0, X_3=0\mid Y=1)\frac{P(Y=1)}{P(X_1=1, X_2=0, X_3=0)} = \underbrace{P(X_1=1\mid \cancel{Y=1})}_{p}\underbrace{P(X_2=\cancel{0}\mid Y=1)}_{q}\underbrace{P(X_3=\cancel{0}\mid X_2=0)}_{1}\frac{\overbrace{P(\cancel{Y=1})}^{1/2}}{P(X_1=1, X_2=0, X_3=0)}$$

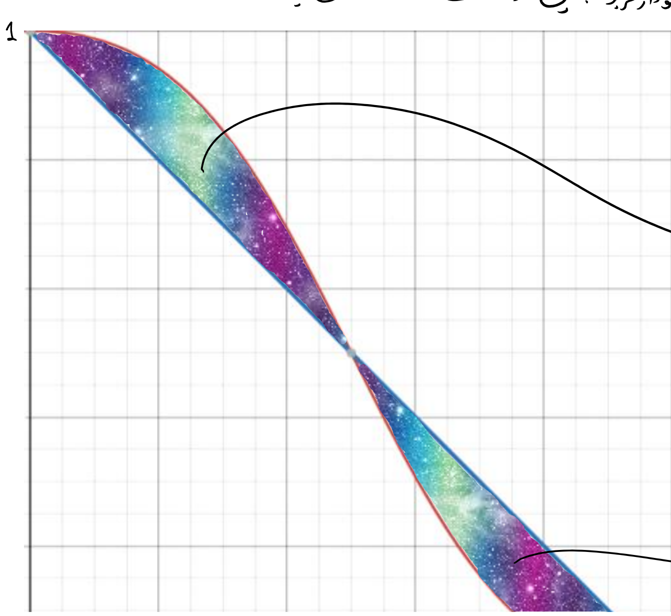
$$= \frac{1}{2}pq\frac{1}{P(X_1=1, X_2=0, X_3=0, Y=0) + P(X_1=1, X_2=0, X_3=0, Y=1)} = \frac{1}{2}pq\frac{1}{P(Y=0)P(X_1=1, X_2=0, X_3=0\mid Y=0) + P(Y=1)P(X_1=1, X_2=0, X_3=0\mid Y=1)}$$

$$= \frac{1}{2}pq\frac{1}{\underbrace{P(\cancel{Y=0})}_{1/2}\underbrace{P(X_1=1\mid \cancel{Y=0})}_{1-p}\underbrace{P(X_2=0\mid \cancel{Y=0})}_{1-q}\underbrace{P(X_3=0\mid \cancel{X_2=0})}_{1} + \underbrace{P(\cancel{Y=1})}_{1/2}\underbrace{P(X_1=1\mid \cancel{Y=1})}_{p}\underbrace{P(X_2=0\mid \cancel{Y=1})}_{q}\underbrace{P(X_3=0\mid \cancel{X_2=0})}_{1}}$$

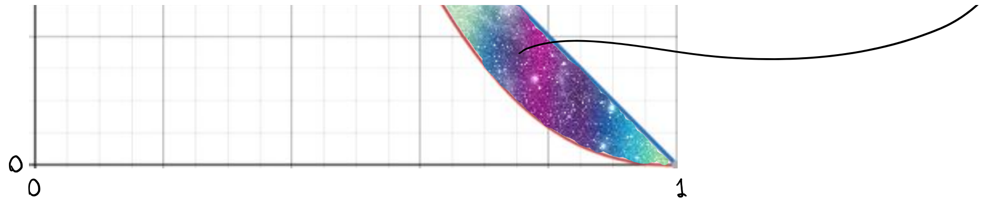
$$= \frac{pq}{(1-p)(1-q) + pq}$$

ج) در دو مدل اگر $P(Y=0, X_1=1, X_2=0, X_3=0) < P(Y=1, X_1=1, X_2=0, X_3=0)$ باشد (در مدل Naive bayes $(1-p)(1-q)^2 < pq^2$ و در مدل دوم اگر $(1-p)(1-q) < pq$ باشند) برچسب برابر با 1 می‌شود. در شکل زیر نمودار مربوط به این دو حالت را مشاهده می‌کنید:

نمودار آبی مربوط به مدل دوم و نمودار قرمز مربوط به مدل اول است.



در این دو ناحیه برچسب ها تفاوت خواهند برد.


0
0
1

الف) این شبکهٔ عصبی به صورت زیر عمل می‌کند:

1- دو نورون بالایی بررسی می‌کنند که آیا x بین اعداد 4 و -4 قرار دارد یا خیر:

1^{st} neuron checks if $x \leq 4$
2^{nd} neuron checks if $-x \leq 4$ (in other words $x \geq -4$)

2- دو نورون پایینی بررسی می‌کنند که آیا y بین اعداد 4 و -4 قرار دارد یا خیر:

3^{rd} neuron checks if $y \leq 4$
4^{th} neuron checks if $-y \leq 4$ (in other words $y \geq -4$)

درصورت برقرار بودن شروط، خروجی step ها 1 خواهد بود. درنتیجه اگر تمامی شروط برقرار باشند خروجی تمامی stepها برابر با 1 خواهد بود و درنتیجه با جمع خروجی حاصل ضرب این اعداد در وزنشان با bias نورون نهایی (-7) خروجی این شبکه برابر با 1 خواهد بود و درغیراینصورت خروجی شبکه 0 است.

بنابراین این شبکه بررسی می‌کند که آیا نقطهٔ ورودی (x,y) درون مربعی 8 در 8 درمرکز دستگاه مختصات قرار دارد یا خیر.

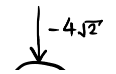
ب) حال که در بخش قبل توانستیم بخش مربعی هدف را پیاده‌سازی کنیم، باید بخش مربوط به لوزی هدف را بسازیم. این لوزی با استفاده از چهار معادله زیر ساخته می‌شود:

$$\begin{cases} y \leq -x + 4\sqrt{2} \\ y \geq -x - 4\sqrt{2} \\ y \leq x + 4\sqrt{2} \\ y \geq x - 4\sqrt{2} \end{cases} \longrightarrow \begin{cases} x + y \leq 4\sqrt{2} \\ x + y \geq -4\sqrt{2} \\ -x + y \leq 4\sqrt{2} \\ -x + y \geq -4\sqrt{2} \end{cases}$$

بنابراین در هرکدام از stepها یک شرطی را بررسی می‌کنیم:

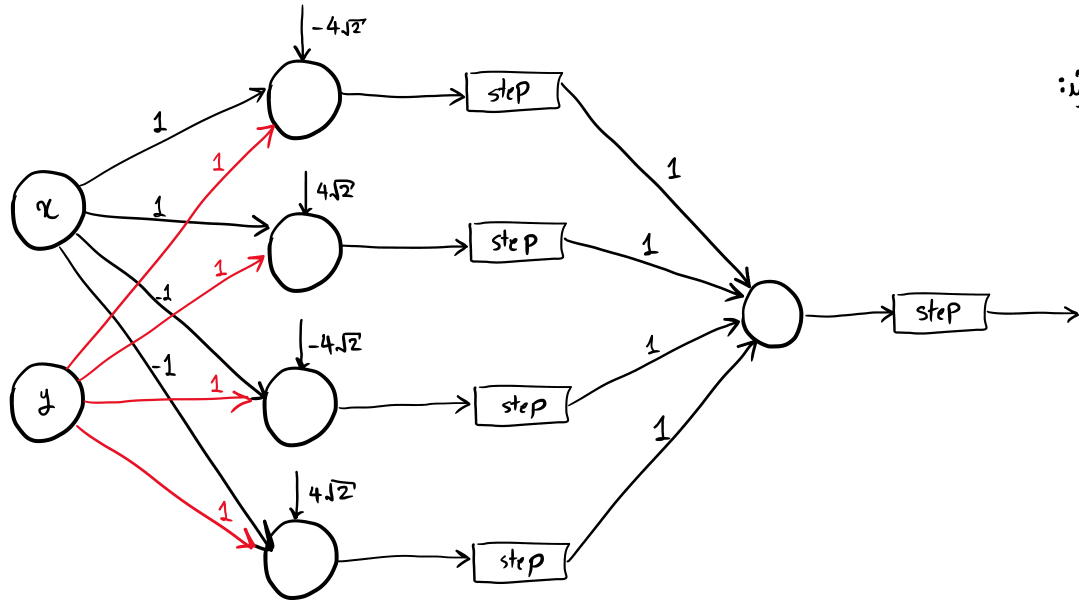
1^{st} neuron: calculate $\underline{x+y-4\sqrt{2}}$, then check if it's lower than zero (return 1 else return 0)
2^{nd} neuron: calculate $\underline{x+y+4\sqrt{2}}$, then check if it's greater than zero (return 1 else return 0)
3^{rd} neuron: calculate $\underline{-x+y-4\sqrt{2}}$, then check if it's lower than zero (return 1 else return 0)
4^{th} neuron: calculate $\underline{-x+y+4\sqrt{2}}$, then check if it's greater than zero (return 1 else return 0)

یک نورون هم برای محاسبهٔ خروجی نهایی می‌سازیم که خروجی این 4 نورون را با هم AND می‌کند. بنابراین درصورت برقراری همهٔ شروط خروجی این شبکه هم 1 خواهد بود.

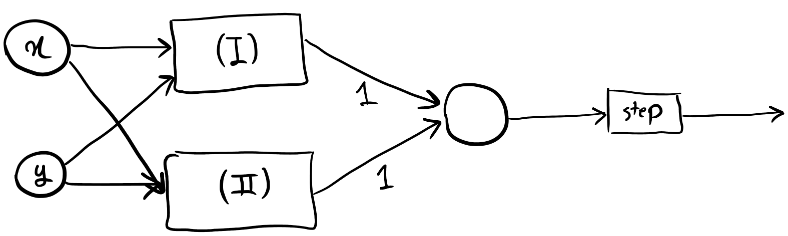


$\underline{1}$ خواهد بود.

این شبکه را مشاهده می کنید:



در نهایت شبکه ای می سازیم که بررسی کند که آیا گلوله به هدف خورده است یا خیر. دو شبکه ای که مثلث و پارسا ساخته اند را در نظر بگیرید (شبکه های (I) و (II)). خروجی های این دو شبکه را در نظر بگیرید؛ اگر گلوله در هر کدام از این محدوده ها برخورد کرده باشد یعنی به هدفمان رسیده ایم. بنابراین یک نورون برای محاسبهٔ جواب نهایی می سازیم که خروجی این دو شبکه را باهم OR کند و خروجی دهد.



ج) می توان با دنبال کردن همین الگو، با ساختن مربع های مشابه با زاویه های متفاوت به دایره رسید (اگر تعداد این مربع ها به بینهایت میل کند یک دایره کامل بی نقص خواهیم داشت.)

الف)

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \quad X = \begin{bmatrix} — 1 — \\ — x_1 — \\ \vdots \\ — x_n — \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\longrightarrow L = \sum_{i=1}^{n} (\beta x_i - y_i)^2 = \| X\beta - Y \|^2 \longrightarrow L = \| X\beta - Y \|^2 = (X\beta - Y)^T (X\beta - Y)$$

ب) از تابع Loss نسبت به β مشتق می‌گیریم:

$$L = (X\beta - Y)^T (X\beta - Y) = (\beta^T X^T - Y^T)(X\beta - Y) = \beta^T X^T X \beta \underbrace{- \beta^T X^T Y - Y^T X \beta}_{-2\beta^T X^T Y} - Y^T Y = \beta^T X^T X \beta - 2\beta^T X^T Y - Y^T Y$$

$$\longrightarrow \frac{\partial L}{\partial \beta} = 2X^T X \beta - 2X^T Y \xrightarrow{\left.\frac{\partial L}{\partial \beta}\right|_{\hat{\beta}} = 0} 2X^T X \hat{\beta} - 2X^T Y = 0 \longrightarrow X^T X \hat{\beta} = X^T Y \longrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

ج)

$$L = \sum_{i=1}^{n} (\beta x_i - y_i)^2 + \lambda \sum_{i=0}^{m} \beta_i^2 = \| X\beta - Y \|^2 + \lambda \| \beta \|^2 = \beta^T X^T X \beta - 2\beta^T X^T Y - Y^T Y + \lambda \beta^T \beta$$

$$\longrightarrow \frac{\partial L}{\partial \beta} = 2X^T X \beta - 2X^T Y + 2\lambda \beta \xrightarrow{\left.\frac{\partial L}{\partial \beta}\right|_{\hat{\beta}} = 0} X^T X \hat{\beta} + \lambda \hat{\beta} = X^T Y \longrightarrow (X^T X + \lambda I)\hat{\beta} = X^T Y$$

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

الف) با توجه به صورت سؤال داریم:

$$B = \min\{\|w\| : \forall i \in [m],\ y_i \langle w, x_i \rangle \geq 1\} = \|w_B\| \longrightarrow y_i w_B^T x_i \geq 1$$

زمان $\underline{t+1}$ را در نظر بگیرید. با استفاده از الگوریتم پرسپترون داریم:

$$w_B^T w^{(t+1)} = w_B^T (w^{(t)} + y_i x_i) = w_B^T w^{(t)} + \underbrace{\boxed{y_i w_B^T x_i}}_{\geq 1} \geq w_B^T w^{(t)} + 1 \geq w_B^T w^{(t-1)} + 2 \geq \cdots \geq w_B^T \underbrace{w^{(1)}}_{0} + t$$

$$\longrightarrow w_B^T w^{(t+1)} \geq t$$

$$\|w^{(t+1)}\|^2 = \|w^{(t)} + y_i x_i\|^2 \leq \|w^{(t)}\|^2 + \|y_i x_i\|^2 \leq \|w^{(t)}\|^2 + R^2 \leq \|w^{(t-1)}\|^2 + 2R^2 \leq \cdots \leq \|\underbrace{w^{(1)}}_{0}\|^2 + tR^2$$

$$\longrightarrow \|w^{(t+1)}\|^2 \leq tR^2 \longrightarrow \|w^{(t+1)}\| \leq \sqrt{t} R$$

$$\longrightarrow t \leq |w_B^T w^{(t+1)}| \leq B\sqrt{t} R \longrightarrow t \leq B\sqrt{t} R \longrightarrow \sqrt{t} \leq BR \longrightarrow t \leq (BR)^2$$

$$,\ t_{max} = T \longrightarrow T = (BR)^2$$

ب) با توجه به راهنمایی سؤال داریم: $\forall i : x_i = e_i$

فرض کنید داریم: $\underline{y_i = 1}$, $\underline{w^* = \sum_{j=1}^{m} e_j}$.

حال برقرار بودن یا نبودن شروط را بررسی می کنیم:

(I) $R = \max \|x_i\| = \max \|e_i\| = 1 \leq 1 \longrightarrow$ first constraint is satisfied

(II) $\|w^*\|^2 = \|\sum_{j=1}^{m} e_j\|^2 = m^2$, $\forall i \in [m] : y_i \langle w^*, x_i \rangle = \langle w^*, e_i \rangle = \langle \sum_{j=1}^{m} e_j, e_i \rangle = \sum_{j=1}^{m} \langle e_j, e_i \rangle$

$$= \underbrace{\langle e_1, e_i \rangle}_{0} + \underbrace{\cdots}_{0} + \underbrace{\langle e_i, e_i \rangle}_{1} + \underbrace{\cdots}_{0} + \underbrace{\langle e_n, e_i \rangle}_{0} = 1 \geq 1 \longrightarrow$$ second constraint is satisfied

(III) $w^{(t+1)} = w^{(t)} + y_i x_i = w^{(t)} + e_i \longrightarrow w^{(t+1)} = \sum_{i=1}^{t} e_i$

فرض کنید الگوریتم تا اینجا ادامه پیدا می کند. بنابراین داریم:

$$\forall i : y_i \langle w^{(t+1)}, x_i \rangle > 0 \longrightarrow \langle \sum_{j=1}^{t} e_j, e_i \rangle > 0 \longrightarrow |\{e_{1:t}\}| \geq m$$

$\longrightarrow t \geq m \longrightarrow$ تعداد m به روزرسانی می شود $\longrightarrow$ third constraint is satisfied