به نام خداوند بخشنده مهربان

هوش مصنوعی

دانشکده مهندسی کامپیوتر

دکتر رهبان
بهار ۱۴۰۳

مهدی علی نژاد، ۴۰۱۱۰۶۲۶۶

تمرین تئوری چهارم

سوال ۱

۱. (۲۰ نمره، درجه سختی ۸) یک دانشجو که ددلاین تمرین ششم درس‌های یادگیری ماشین و معماری کامپیوترش در یک شب قرار گرفته و فردای آن نیز امتحان درس دیگری را دارد، می‌خواهد با توجه به سه پارامتر تعداد روزهای معین شده برای تحویل تکلیف، درجه سختی تکلیف و کسری از کلاس که تا یک روز قبل از ددلاین تمرین را تحویل داده‌اند، پیش‌بینی کند که کدام تمرین تمدید می‌شود. او می‌خواهد این پیش‌بینی را به کمک درخت تصمیم و با استفاده از داده‌هایی که از پنج تمرین قبلی این دو درس دارد، انجام دهد. اگر هر کدام از این سه پارامتر را به ترتیب با x_1، x_2 و x_3 نشان دهیم، داده‌هایی که از ۵ تمرین قبل داریم و همچنین اطلاعات تمرین ششم (که باید خروجی آن را پیش‌بینی کنیم) در جدول زیر دیده می‌شوند:

Machine Learning					Computer Architecture				
HW	x_1	x_2	x_3	y	HW	x_1	x_2	x_3	y
1	6	1	0.4	1	1	7	1	0.3	1
2	8	1	0.6	1	2	6	0	0.7	0
3	7	0	0.6	0	3	8	1	0.1	0
4	5	0	0.5	1	4	5	0	0.4	0
5	8	1	0.5	0	5	7	0	0.1	1
6	7	0	0.55	?	6	8	1	0.4	?

با استفاده از شاخص Information Gain درخت تصمیم را برای هر کدام از درس‌ها بدست آورید و تخمین بزنید که تمرین ششم هر درس تمدید می‌شود یا خیر و از آن نتیجه بگیرید که دانشجو باید کدام تمرین را زودتر شروع کند. برای ساده‌تر شدن محاسبات، بعد از استفاده از هر متغیر در یک گره، آن را کنار گذاشته و در گره‌های بعدی از آن استفاده نکنید. (هر جا حالت $IG(x_i) = IG(x_j)$ پیش آمد که در آن $i < j$، متغیر x_i را انتخاب کنید.)

برای تشکیل درخت تصمیم، نیاز است تا entropy مختص به y طبق فرمول $\mathrm{H}(S) = -\sum_{i=1}^{k} \mathrm{P}(S = s_i) \log_2 \mathrm{P}(S = s_i)$ و y به شرط تمام x_i ها را بدست آوریم با استفاده از فرمول $\mathrm{H}(S|V) = -\sum_{j=1}^{m} \mathrm{P}(V = v_j) \sum_{i=1}^{k} \mathrm{P}(S = s_i|V = v_j) \log_2 \mathrm{P}(S = s_i|V = v_j)$ و سپس متغیری که بیشترین information gain را به ما می دهد طبق فرمول $\mathrm{IG}(S, V) = \mathrm{H}(S) - \mathrm{H}(S|V)$ را در راس تصمیم گیری قرار دهیم و به طور بازگشتی به سمت پایین حرکت کنیم تا تمام اطلاعات دسته بندی شوند.
برای درس ML داریم:

$$\mathrm{H}(Y) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$$

$$\mathrm{H}(Y|X_1) = [\text{for} \quad j \quad \text{in } \{6,7,8\} : -\mathrm{P}(X_1 >= j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_1 >= j)\log_2 \mathrm{P}(\mathrm{Y = i|X_1 >= j})$$

$$-\,\mathrm{P}(X_1 < j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_1 < j)\log_2 \mathrm{P}(\mathrm{Y = i|X_1 < j})] = [0.8, 0.55, 0.95]$$

$$\mathrm{H}(Y|X_2) = -\sum_{j=0}^{1}\mathrm{P}(X_2 = j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_2 = j)\log_2 \mathrm{P}(\mathrm{Y = i|X_2 = j}) = 0.4 + 0.55 = 0.95$$

$$\mathrm{H}(Y|X_3) = [\text{for} \quad j \quad \text{in } \{0.5, 0.6\} : -\mathrm{P}(X_1 >= j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_1 >= j)\log_2 \mathrm{P}(\mathrm{Y = i|X_1 >= j})$$

$$-\,\mathrm{P}(X_1 < j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_1 < j)\log_2 \mathrm{P}(\mathrm{Y = i|X_1 < j})] = [0.8, 0.95]$$

$$\mathrm{IG}(Y, X_1) = [0.17, 0.42, 0.02]$$

$$\mathrm{IG}(Y, X_2) = 0.02$$

$$\mathrm{IG}(Y, X_3) = [0.17, 0.02]$$

پس X_1 را در راس قرار می دهیم با مرز جدایی ۷ ، سپس در شاخه ی کوچکتر مساوی، خلوص صد در صدی داریم پس نیازی به دسته بندی بیشتر نیست ولی در طرف بزرگتر داریم:

$$\mathrm{H}(Y) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.91$$

$$\mathrm{H}(Y|X_2) = -\sum_{j=0}^{1}\mathrm{P}(X_2 = j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_2 = j)\log_2 \mathrm{P}(\mathrm{Y = i|X_2 = j}) = 0.66$$

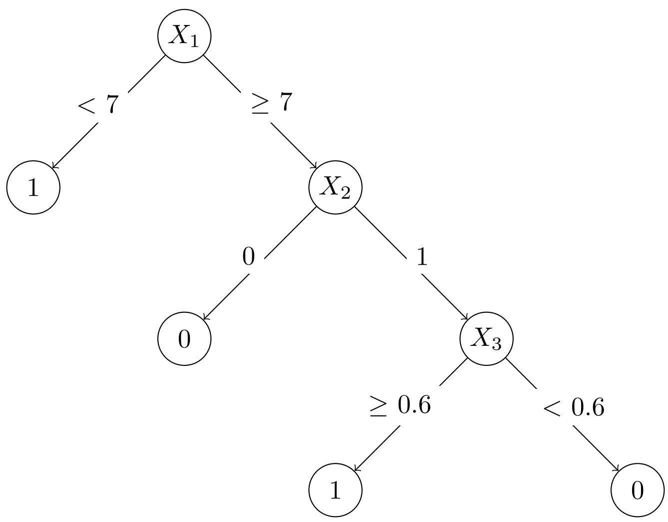
$$\mathrm{H}(Y|X_3) = [\text{for} \quad j \quad \text{in } \{0.6\} : -\mathrm{P}(X_1 >= j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_1 >= j)\log_2 \mathrm{P}(\mathrm{Y = i|X_1 >= j})$$

$$-\,\mathrm{P}(X_1 < j)\sum_{i=0}^{1}\mathrm{P}(Y = i|X_1 < j)\log_2 \mathrm{P}(\mathrm{Y = i|X_1 < j})] = [0.66]$$

$$\mathrm{IG}(Y, X_2) = 0.25$$

$$\mathrm{IG}(Y, X_3) = [0.25]$$

در این سطح نیز X_2 را راس قرار می دهیم و در سمتی که همچنان خالص نیست نیز، طبق متغیر باقی مانده جدا سازی می کنیم و درخت نهایی مان برابر است با:



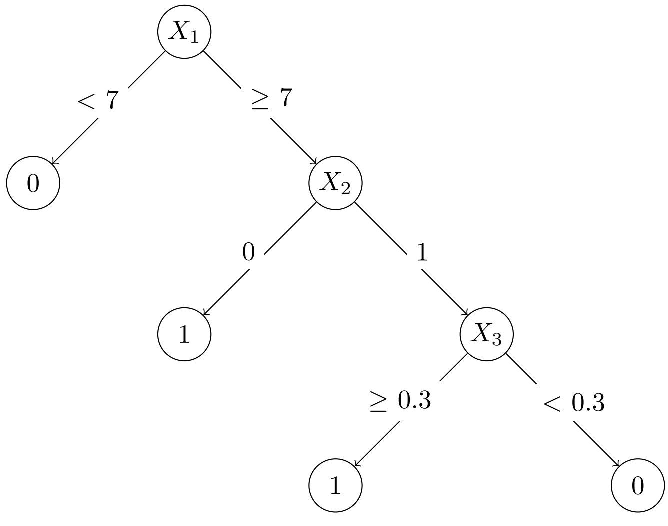
برای درس CA داریم:

$$
\begin{aligned}
&\mathrm{H}(Y) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97\\
&\mathrm{H}(Y|X_1) = [\text{for} \quad j \quad \text{in } \{6,7,8\} : -\mathrm{P}(X_1 >= j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_1>=j)\log_2\mathrm{P(Y=i|X_1>=j)}\\
&-\mathrm{P}(X_1<j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_1<j)\log_2\mathrm{P(Y=i|X_1<j)}] = [0.8, 0.55, 0.55]\\
&\mathrm{H}(Y|X_2) = -\sum_{j=0}^{1}\mathrm{P}(X_2=j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_2=j)\log_2\mathrm{P(Y=i|X_2=j)} = 0.4+0.55 = 0.95\\
&\mathrm{H}(Y|X_3) = [\text{for} \quad j \quad \text{in } \{0.3,0.4,0.7\} : -\mathrm{P}(X_1 >= j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_1>=j)\log_2\mathrm{P(Y=i|X_1>=j)}\\
&-\mathrm{P}(X_1<j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_1<j)\log_2\mathrm{P(Y=i|X_1<j)}] = [0.95, 0.55, 0.8]\\
&\mathrm{IG}(Y,X_1) = [0.17, 0.42, 0.42]\\
&\mathrm{IG}(Y,X_2) = 0.02\\
&\mathrm{IG}(Y,X_3) = [0.02, 0.42, 0.17]
\end{aligned}
$$

پس X_1 را در راس قرار می دهیم با مرز جدایی ۷ ، سپس در شاخه ی کوچکتر مساوی، خلوص صد در صدی داریم پس نیازی به دسته بندی بیشتر نیست ولی در طرف بزرگتر داریم:

$$
\begin{aligned}
&\mathrm{H}(Y) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.91\\
&\mathrm{H}(Y|X_2) = -\sum_{j=0}^{1}\mathrm{P}(X_2=j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_2=j)\log_2\mathrm{P(Y=i|X_2=j)} = 0.66\\
&\mathrm{H}(Y|X_3) = [\text{for} \quad j \quad \text{in } \{0.3\} : -\mathrm{P}(X_1 >= j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_1>=j)\log_2\mathrm{P(Y=i|X_1>=j)}\\
&-\mathrm{P}(X_1<j)\sum_{i=0}^{1}\mathrm{P}(Y=i|X_1<j)\log_2\mathrm{P(Y=i|X_1<j)}] = [0.66]\\
&\mathrm{IG}(Y,X_2) = 0.25\\
&\mathrm{IG}(Y,X_3) = [0.25]
\end{aligned}
$$

در این سطح نیز X_2 را راس قرار می دهیم و در سمتی که همچنان خالص نیست نیز، طبق متغیر باقی مانده جدا سازی می کنیم و درخت نهایی مان برابر است با:



طبق این دو درخت تخمین زده شده از اطلاعات، تمرین ML او تمدید نمی شود ولی تمرین CA او تمدید می شود.

سوال ۲

۲. (۱۵ نمره، درجه سختی ۸) فرض کنید یک مسئله‌ی دسته‌بندی[1] داریم که درآن برچسب[2] آن (Y) و فیچر های آن ($X_۱, X_۲, X_۳$)، متغیّرهای بولی هستند. فیچرهای $X_۱$ و $X_۲$ به شرط Y مستقل هستند و همچنین فیچر $X_۳$، یک فیچر تکراری از $X_۲$ است. ($X_۲ = X_۳$)
اگر بدانیم:

$$\mathbb{P}(X_۱ = ۱|Y = ۱) = p$$
$$\mathbb{P}(X_۱ = ۱|Y = ۰) = ۱ - p$$
$$\mathbb{P}(X_۲ = ۰|Y = ۱) = q$$
$$\mathbb{P}(X_۲ = ۰|Y = ۰) = ۱ - q$$
$$\mathbb{P}(Y = ۱) = \mathbb{P}(Y = ۰) = ۰/۵$$

حال فرض کنید که یک داده تست بدون برچسب به شرح زیر به ما داده شده است:

$$X_۱ = ۱, X_۲ = X_۳ = ۰$$

حال می‌خواهیم با پیش‌بینی Y، این داده را طبقه‌بندی کنیم.

(آ) با فرض مدل بیز ساده‌لوحانه، قاعده تصمیم‌گیری را به ازای $Y = ۱$ و برحسب p و q بیابید.

(ب) بدون فرض مدل بیز ساده‌لوحانه، قاعده تصمیم‌گیری بهینه را به ازای $Y = ۱$ و برحسب p و q بیابید.

(ج) با فرض آنکه محور افقی q و محور عمودی p باشد، مرز تصمیم‌گیری (آ) و (ب) را در بازه‌ی (۰, ۱) رسم کرده و مشخص کنید در چه نواحی قاعده‌ی تصمیم‌گیری مدل بیزساده‌لوحانه نسبت به قاعده‌ی تصمیم‌گیری بهینه دچار خطا می‌شود.

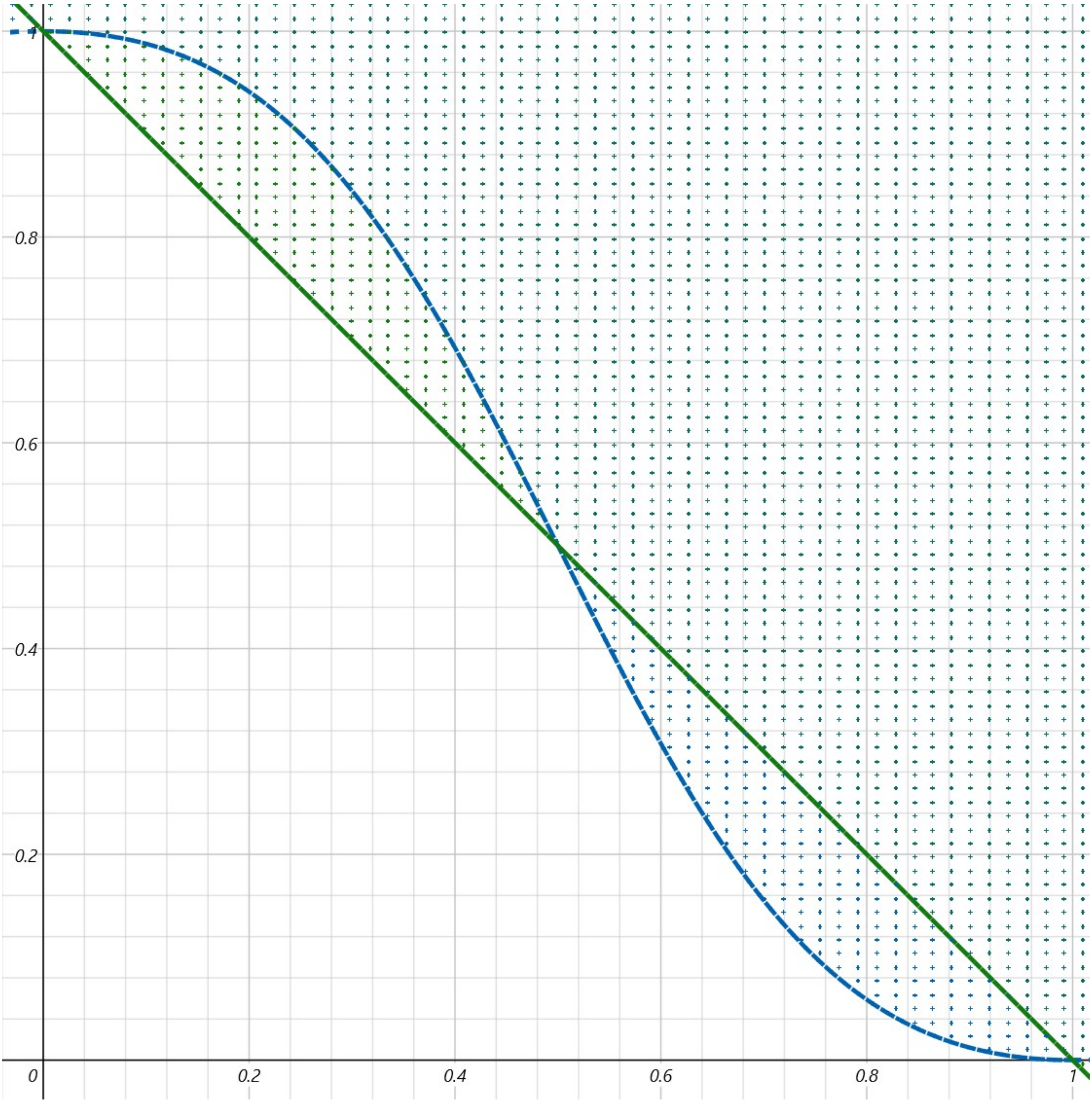
(آ)

$$\mathrm{P}(Y=1|X_1=1,X_2=0,X_3=0) = \mathrm{P}(Y=1)\prod \mathrm{P}(X_i|Y=1) = \frac{1}{2}\times p\times q\times q = \frac{pq^2}{2}$$
$$\mathrm{P}(Y=0|X_1=1,X_2=0,X_3=0) = \mathrm{P}(Y=0)\prod \mathrm{P}(X_i|Y=0) = \frac{1}{2}\times(1-p)\times(1-q)\times(1-q)$$
$$= \frac{1-p-q+pq-q+pq+q^2-pq^2}{2} = \frac{1-p-2q+2pq+q^2-pq^2}{2}$$
$$\mathrm{P}(Y=1|X_1=1,X_2=0,X_3=0) > \mathrm{P}(Y=0|X_1=1,X_2=0,X_3=0) \to pq^2 > 1-p-2q+2pq+q^2-pq^2$$
$$\to 2pq^2+p+2q > 1+2pq+q^2 \to p > \frac{1+q^2-2q}{2q^2+1-2q}$$

(ب)

$$\mathrm{P}(X_1=1,X_2=X_3=0|Y=1) \overset{independent}{=} \mathrm{P}(X_1=1|Y=1)\mathrm{P}(X_2=0|Y=1) = pq$$
$$\mathrm{P}(X_1=1,X_2=X_3=0|Y=0) \overset{independent}{=} \mathrm{P}(X_1=1|Y=0)\mathrm{P}(X_2=0|Y=0) = (1-p)(1-q)$$
$$\mathrm{P}(Y=1|X_1=1,X_2=0,X_3=0) > \mathrm{P}(Y=0|X_1=1,X_2=0,X_3=0) \to pq > 1-p-q+pq \to p+q > 1$$

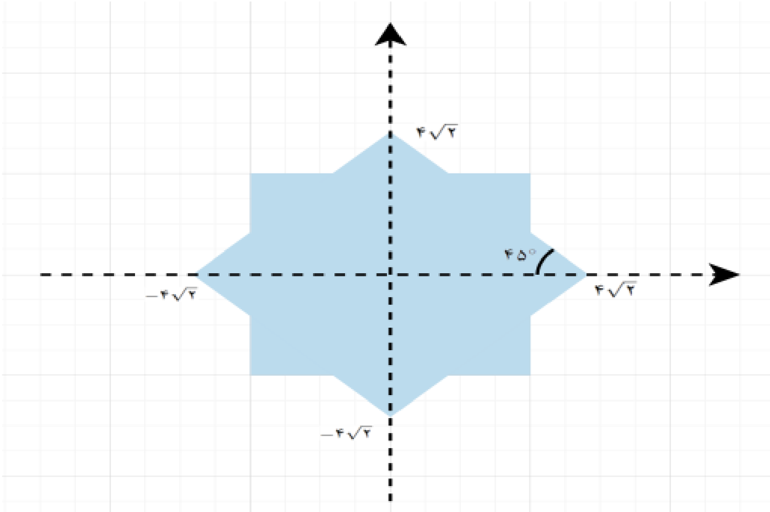
(ج) .



خط سبز نشان دهنده ی مرز واقعی بین این label هاست و خط آبی مدل بیز ساده لوحانه است. مدل بیز ساده لوحانه در نواحی بین این دو خط دچار خطا می شود. نمودار p برحسب q است.

۳. (۱۵ نمره، درجه سختی ۶) در یک روز پائیزی، پارسا و متین در حال سپری کردن زمان استراحت بین کلاس‌هایشان بر روی چمن های اکلیلی[۳] هستند که به یکباره بحث تیراندازی به میان کشیده می‌شود! متین ناگهان ایده‌ای به ذهنش خطور می‌کند و پس از مطرح کردن آن با پارسا، با جمله‌ی همیشگی "آقاااا برییییم پیپرش کنیم!" روبه‌رو می‌شود.

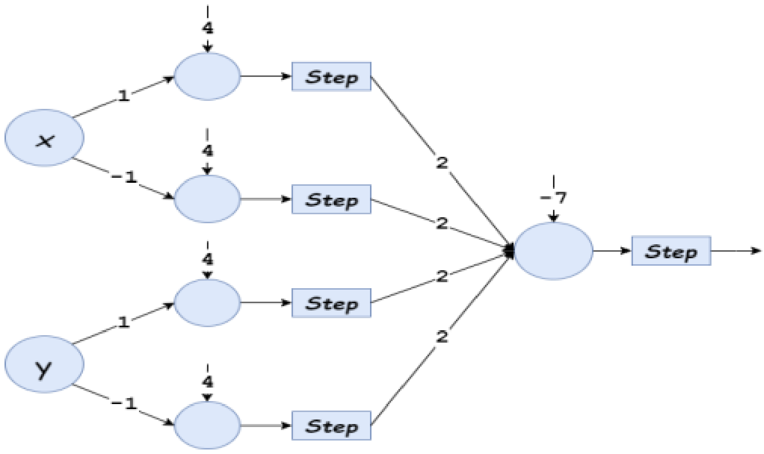
ایده‌ی متین این است که بتوانیم به کمک یک شبکه عصبی، تشخیص دهیم آیا تیر به محدوده‌ی هدف خورده‌است یا خیر. متین ابتدا مدل ساده‌لوحانه شکل ۱ از هدف را پیشنهاد می‌کند:



شکل ۱: خروجی مدل ساده‌لوحانه هدف (هرخانه یک واحد است و $\sqrt{۲} \approx ۱/۴$)

حال به بخش های زیر پاسخ دهید.(در ابتدا فقط به تابع فعال‌ساز Step دسترسی داریم.)

(آ) متین و پارسا تصمیم می‌گیرند طراحی این شبکه عصبی را بین خودشان تقسیم کنند، بنابراین پارسا شبکه شکل ۲ را طراحی کرده است. خروجی شبکه‌ای که او طراحی کرده است، را مشخص کنید.



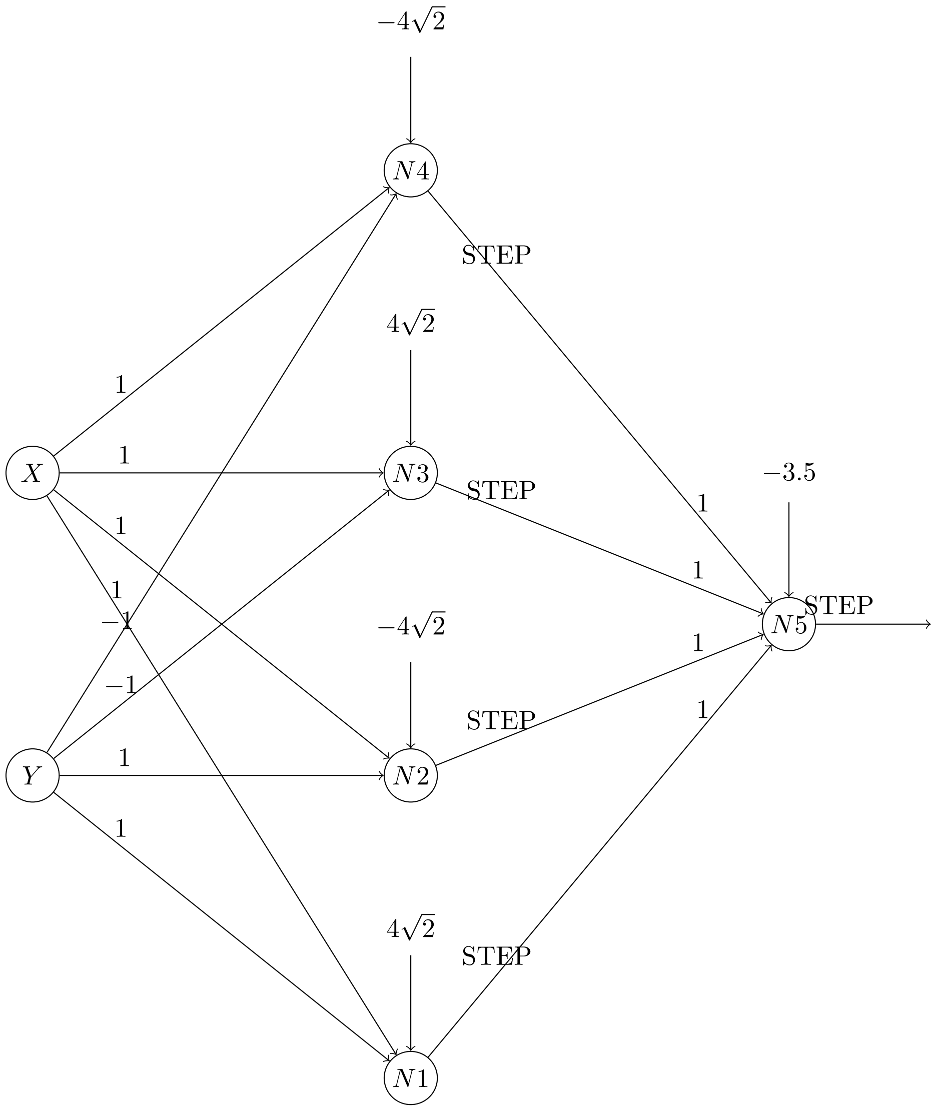
شکل ۲: شبکه عصبی طراحی شده توسط پارسا

(ب) حال متین باید باقی شبکه را به کمک طراحی پارسا کامل کند تا به خروجی شکل ۱ برسند. در این خصوص به متین کمک کنید و شبکه مورد نظر را به کمک بخش قبلی طراحی کنید.

(ج) اما خروجی شکل ۱ از شبکه طراحی‌شده همانطور که مشخص است خیلی ساده‌لوحانه است! زیرا از یک هدف ایده‌آل انتظار می‌رود، شکل و شمایل دایروی داشته باشد. آیا می توانیم چنین دایره بی نقصی طراحی کنیم؟ به کمک روندی که در بخش های قبلی دنبال کردیم، راهی پیشنهاد دهید که مدلی واقع‌گرایانه‌تر از هدف بتوانیم طراحی کنیم.

(آ) این شبکه ی عصبی، بودن نقطه ی مورد نظر در داخل مربعی ۴ در ۴ بر روی مرکز مختصات را بررسی می کند، به این صورت که پروسپترون اول چک می کند که x بزرگتر از ۴− باشد سپس پروسپترون دوم چک می کند که کمتر از ۴ نیز باشد و دو نورون بعدی نیز مسئول چک کردن همین شروط برای y هستند، و سپس تمام این نرون ها با هم and می شوند.

(ب) تا شکل مورد نظر یک مربع دیگر فاصله داریم که به صورت اورب و احاطه شده توسط خط های $x+y=-۴\sqrt{۲}, x+y=۴\sqrt{۲}, x-y=-۴\sqrt{۲}, x-y=۴\sqrt{۲}$ است. شبکه ی عصبی آن را به این صورت طراحی می کنیم.



سپس کافیست خروجی این شبکه عصبی را با وزن یک با خروجی شبکه عصبی پارسا با وزن یک ترکیب کنیم و مقدار b را برابر ۰/۵− قرار دهیم.

(ج) می توانیم با دنبال کردن همین الگو و افزایش تعداد مربع ها، هرچه تعداد مربع ها بیشتر شوند، به شکل دایره نزدیک تر می شویم. زاویه ی خط های تشکیل دهنده ی این مربع ها نیز اگر n تعداد مربع ها باشد، $\frac{\pi}{۲n}i \quad \text{for} \quad i \in \{۱, ..., n\}$

سوال ۴

۴. (۲۰ نمره، درجه سختی ۶) همانطور که در کلاس درس یاد گرفته‌اید. مسئله رگرسیون خطی به شکل زیر تعریف می‌شود:

$$y_i = \sum_{j=۱}^{m} \beta_j x_{ij} + \beta.$$

که در آن هدف مینیمم کردن تابع لاس است: $L = \sum_{i=۱}^{n} (\beta \cdot x_i - y_i)^۲$

الف) مسئله رگرسیون خطی را به شکل ماتریسی بنویسید.

ب) اثبات کنید جواب این مسئله برابر است با: $\hat{\beta} = (X^T X)^{-۱} X^T y$

ج) در مسئله Ridge Regression برای اینکه ضرایب مقدار زیادی نداشته باشند به تابع لاس یک جمله به شکل $\sum \beta_i^۲$ اضافه می‌شود که در نهایت تابع لاس به شکل زیر تبدیل می‌شود:

$$L = \sum_{i=۱}^{n} (\beta \cdot x_i - y_i)^۲ + \lambda \sum_{i=۰}^{m} \beta_i^۲$$

برای این مسئله اثبات کنید که جواب به شکل زیر بدست می‌آید:

$$\hat{\beta} = (X^T X + \lambda I)^{-۱} X^T y$$

(آ)

$$y_i = \sum_{j=1}^{m} \beta_j x_{ij} + \beta_0 = [\beta_1, \beta_2, \dots, \beta_m] \begin{bmatrix} x_{i1} \\ x_{i2} \\ . \\ . \\ . \\ x_{im} \end{bmatrix} + \beta_0 = [\beta_0, \beta_1, \dots, \beta_m] \begin{bmatrix} 1 \\ x_{i1} \\ . \\ . \\ . \\ x_{im} \end{bmatrix} = \beta^T x_i$$

$$L = \sum_{i=1}^{n} (\beta^T . x_i - y_i)^2 = \left\| [\beta_0, \beta_1, \dots, \beta_m] . \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ x_{1m} & x_{2m} & \dots & x_{nm} \end{bmatrix} - [y_1, y_2, \dots, y_n] \right\|_2 = (\beta^T \times X - Y).(\beta^T \times X - Y)$$

(ب) برای این منظور مشتق می گیریم و برابر صفر قرار می دهیم.

$$L = \beta^T X X^T \beta + Y Y^T - 2\beta^T X Y^T$$
$$\frac{\partial L}{\partial \beta} = 2X^T X \beta - 2X^T Y \xrightarrow{=0} X^T X \hat{\beta} = X^T Y \rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

(ج) مانند حالت ب عمل می کنیم:

$$L' = L + \lambda ||\beta||_2$$
$$\frac{\partial L'}{\partial \beta} = \frac{\partial L}{\partial \beta} + \frac{\partial \lambda \beta^T \beta}{\partial \beta} = 2X^T X \beta - 2X^T Y + 2\lambda \beta \xrightarrow{=0} (X^T X - \lambda I)\hat{\beta} = X^T Y \rightarrow \hat{\beta} = (X^T X - \lambda I)^{-1} X^T Y$$

۵. (۲۰ نمره، درجه سختی ۶)

* نام‌گذاری خاص: $[n]$: the set $\{1, ..., n\}$ (for $n \in \mathbb{N}$)

در این سوال می‌خواهیم به بررسی الگوریتم پرسپترون برای دسته‌بندی[۴] نیم‌فضاها[۵] بپردازیم:

(آ) از اسلاید های درس به خاطر دارید که الگوریتم پرسپترون(یادگیری پرسپترون)، یک الگوریتم بازگشتی[۶] است که با دریافت یک مجموعه آموزش[۷]، طیّ T تکرار[۸] دنباله‌ای از بردار های وزنِ $\{\mathbf{w}^{(t)}\}_{t=۱}^{T}$ را می‌سازد. (شکل۱)

Batch Perceptron

input: A training set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$
initialize: $\mathbf{w}^{(1)} = (0, \ldots, 0)$
for $t = 1, 2, \ldots$
  if $(\exists\, i \text{ s.t. } y_i\langle \mathbf{w}^{(t)}, \mathbf{x}_i\rangle \le 0)$ **then**
    $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + y_i\mathbf{x}_i$
  else
    output $\mathbf{w}^{(t)}$

شکل ۳: الگوریتم پرسپترون

حال فرض کنید، که $(\mathbf{x}_۱, y_۱), ..., (\mathbf{x}_m, y_m)$ تفکیک پذیر هستند و همچنین اگر داشته باشیم:

$$B = \min\{\|\mathbf{w}\| : \forall i \in [m], y_i\langle \mathbf{w}, \mathbf{x}_i\rangle \ge ۱\}$$
$$R = \max_i \|\mathbf{x}_i\|$$

نشان دهید آنگاه الگوریتم پرسپترون پس از پیمودن حداکثر $(RB)^۲$ تکرار متوقف می‌شود و هر زمان که متوقف شود، خواهیم داشت:

$$\forall i \in [m], y_i\langle \mathbf{w}^{(t)}, \mathbf{x}_i\rangle > ۰$$

(ب) نشان دهید در حالت زیر نیز حکم قسمت (الف) برقرار است:

برای هر عدد صحیح مثبت m، یک بردار $\mathbf{w}^* \in \mathbb{R}^d$ و یک دنباله داده $\{(\mathbf{x}_i, y_i)\}_{i=۱}^{m}$ وجود دارد بگونه ای که شرایط زیر برقرار باشد:

- $R = \max_i \|\mathbf{x}_i\| \le ۱$
- $\|\mathbf{w}^*\|^۲ = m$ ، $\forall i \in [m], y_i\langle \mathbf{w}^*, \mathbf{x}_i\rangle \ge ۱$

دقت کنید که به کمک نماد های قسمت (الف) خواهیم داشت:

$$B = \min\{\|\mathbf{w}\| : \forall i \in [m], y_i\langle \mathbf{w}, \mathbf{x}_i\rangle \ge ۱\} \le \sqrt{m}$$

- با اجرای الگوریتم پرسپترون روی این دنباله تا قبل از همگرایی، تعداد m بروزرسانی انجام می‌شود.

* راهنمایی: $d = m$ را انتخاب کرده و به‌ازای هر i، $\mathbf{x}_i = \mathbf{e}_i$ اختیار کنید.

(آ) ابتدا بزرگ شدن norm را محاسبه می کنیم. سپس بزرگ شدن ضرب داخلی بردار w با بردار هدف را نظارت می کنیم و در انتها نیز مشخص می کنیم تا چه تعداد می توان این w را آپدیت کرد. فرض می کنیم w^* بردار حاصل B است و تمام نقاط را جدا می کند و آن را بردار هدف فرض می کنیم.

$$
\begin{aligned}
||w^{(t+1)}||^2 &= ||w^{(t} + y_i x_i||^2 = ||w^{(t)}||^2 + ||x_i||^2 + 2y_i\langle w^{(t)}, x_i\rangle \\
&\leq ||w^{(t)}||^2 + ||x_i||^2 \leq ||w^{(t)}||^2 + R^2 \\
&||w^{(t+1)}||^2 \leq tR^2 \\
\langle w^*, w^{(t+1)}\rangle &= \langle w^*, w^{(t)} + y_i x_i\rangle = \langle w^*, w^{(t)}\rangle + \langle w^*, y_i x_i\rangle \\
&\geq \langle w^*, w^{(t)}\rangle + 1 \\
&\langle w^*, w^{(t+1)}\rangle \geq t \\
\langle w^*, w^{(t+1)}\rangle &\xrightarrow{\text{Cauchy–Schwarz inequality}} \leq ||w^*||||w^{(t+1)}|| \\
&t \leq \langle w^*, w^{(t+1)}\rangle \leq ||w^*||||w^{(t+1)}|| = B\sqrt{t}R \\
&\sqrt{t} \leq BR \\
&t \leq (BR)^2
\end{aligned}
$$

پس حداکثر شاهد $(BR)^۲$ تکرار خواهیم بود.

(ب) طبق راهنمایی گفته شده، $d = m$ قرار می دهیم و بردار های X_i را برابر با بردار یکه های واحد e_i قرار می دهیم و $y_i = ۱$. اگر $w^{(۰)} = O$ طبق الگوریتمی که داشتیم می تواند نشان داد که $w^{(i)} = \Sigma_{j<i} e_j$ زیرا در حلقه ی اول حاصل تمامی ضرب داخلی ها صفر است و نیاز است که آپدیت رخ دهد. پس از این حلقه داریم:

$$w = (1, 1, \cdot, 1)$$

که با تمامی داده ها همسانی دارد و آنها را با label آنها، درست match می کند. همچنین norm آن نیز برابر $\sqrt{m}$ است که شرط سوال را براورده می کند.