

سوالات کوتاه پاسخ

1. یک طبقه‌بند بیز ساده را با ۳ متغیر ورودی بولی، X_1 ، X_2 و X_3 ، و یک خروجی بولی، Y ، در نظر بگیرید. • چند پارامتر باید برای آموزش چنین طبقه‌بندی بیز ساده‌ای تخمین زده شوند؟ (نیازی نیست آنها را لیست کنید، فقط تعداد کل را بدهید)

پاسخ‌ها:

برای یک طبقه‌بند بیز ساده، باید، $P(Y=1)$ ، $P(X_1=1|Y=0)$ ، $P(X_2=1|Y=0)$ ، $P(X_3=1|Y=0)$ ، $P(X_1=1|Y=1)$ ، $P(X_2=1|Y=1)$ و $P(X_3=1|Y=1)$ را تخمین بزنیم. سایر احتمالات را می‌توان با استفاده از محدودیت جمع شدن احتمالات به ۱ بدست آورد. بنابراین نیاز به تخمین ۷ پارامتر داریم.

2. چند پارامتر باید تخمین زده شوند تا طبقه‌بند فوق بدون فرض استقلال شرطی بیز ساده یاد گرفته شود؟

پاسخ‌ها:

بدون فرض استقلال شرطی، همچنان نیاز به تخمین $P(Y=1)$ داریم. برای $Y=1$ ، باید تمام حالت‌های (X_1, X_2, X_3) را بدانیم، یعنی 3^2 حالت ممکن برای (X_1, X_2, X_3) . با توجه به محدودیت جمع شدن احتمالات به ۱، باید $3^2 - 1 = 7$ پارامتر برای $Y=1$ تخمین بزنیم. بنابراین تعداد کل پارامترها $1 + (3^2 - 1) = 15$ می‌باشد.

3. (صحیح یا غلط؟) خطای یک فرضیه که بر روی مجموعه آموزشی اندازه‌گیری شده است، یک تخمین بدبینانه از خطای واقعی فرضیه ارائه می‌دهد.

پاسخ‌ها:

غلط. خطای آموزشی به صورت خوش‌بینانه بایاس شده است زیرا معمولاً کمتر از خطای واقعی است.

4. (صحیح یا غلط؟) اگر m نقطه داده داده شود و نیمی از آنها برای آموزش و نیمی برای آزمون استفاده شود، تفاوت بین خطای آموزشی و خطای آزمون با افزایش m کاهش می‌یابد.

پاسخ‌ها:

صحیح. با افزایش تعداد داده‌ها، خطای آموزشی افزایش و خطای آزمون کاهش می‌یابد و هر دو به خطای واقعی همگرا می‌شوند.

5. (صحیح یا غلط؟) بیش برآزش بیشتر زمانی رخ می‌دهد که مجموعه داده‌های آموزشی کوچک باشد.

پاسخ‌ها:

صحیح. با مجموعه داده‌های آموزشی کوچک، راحت‌تر می‌توان یک فرضیه پیدا کرد که داده‌های آموزشی را به طور کامل براز کند، یعنی بیش برآزش کند.

6. (صحیح یا غلط؟) بیش برآزش بیشتر زمانی رخ می‌دهد که فضای فرضیه کوچک باشد.

پاسخ‌ها:

غلط. از توازن بایاس-واریانس می‌توان این را دید. وقتی فضای فرضیه کوچک است، بایاس بیشتر و واریانس کمتر است. بنابراین با یک فضای فرضیه کوچک، کمتر احتمال دارد که یک فرضیه پیدا شود که داده‌ها را به خوبی براز کند، یعنی بیش برآزش کند.

درخت تصمیم

یک دانشجو که ددلاین تمرین 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 Gain Information 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.)

حل. برای درس معماری کامپیوتر:

$$H(y) = - \sum_i P(y = y_i) \log(P(y = y_i)) = -\frac{3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} \approx 0.292$$

$$\text{Threshold } x_1 \leq 5 : H(y|x_1) = -\frac{1}{5}(1 \log 1 + 0 \log 0) - \frac{4}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) \approx 0.24$$

$$\text{Threshold } x_1 \leq 6 : H(y|x_1) = -\frac{2}{5}(1 \log 1 + 0 \log 0) - \frac{3}{5}(\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3}) \approx 0.166$$

$$\text{Threshold } x_1 \leq 7 : H(y|x_1) = -\frac{4}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{5}(1 \log 1 + 0 \log 0) \approx 0.24$$

$$\text{Threshold } x_1 \leq 8 : H(y|x_1) = H(y) \approx 0.292$$

$$\text{Best threshold for } x_1 \text{ is } 6 \implies IG(x_1) \approx 0.126$$

$$H(y|x_2) = -\frac{2}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{3}{5}(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3}) \approx 0.286$$

$$IG(x_2) \approx 0.006$$

$$\text{Threshold } x_3 \leq 0.1 : H(y|x_3) = -\frac{2}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{3}{5}(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3}) \approx 0.286$$

$$\text{Threshold } x_3 \leq 0.3 : H(y|x_3) = -\frac{3}{5}(\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3}) - \frac{2}{5}(1 \log 1 + 0 \log 0) \approx 0.166$$

$$\text{Threshold } x_3 \leq 0.4 : H(y|x_3) = -\frac{4}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{5}(1 \log 1 + 0 \log 0) \approx 0.24$$

$$\text{Best threshold for } x_3 \text{ is } 0.3 \implies IG(x_3) \approx 0.126$$

از آنجایی که $IG(x_1) = IG(x_3)$ شد، یکی را به صورت تصادفی انتخاب می‌کنیم (مثلاً x_1) و به عنوان ریشه درخت تصمیم قرار می‌دهیم. حال بقیه‌ی درخت را می‌سازیم:

$$H(y) = - \sum_i P(y = y_i) \log(P(y = y_i)) = -\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} \approx 0.276$$

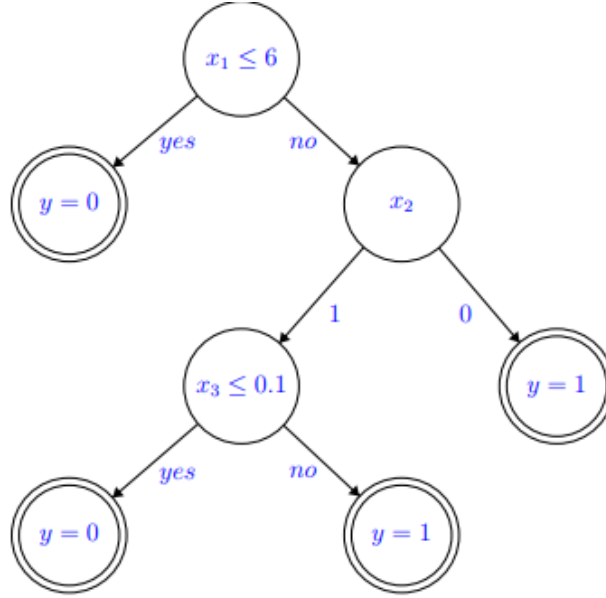
$$H(y|x_2) = -\frac{2}{3}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{3}(1 \log 1 + 0 \log 0) \approx 0.2$$

$$IG(x_2) \approx 0.76$$

$$\text{Threshold } x_3 \leq 0.1 : H(y|x_3) = -\frac{2}{3}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{3}(1 \log 1 + 0 \log 0) \approx 0.2$$

$$IG(x_3) \approx 0.76$$

از آنجایی که $IG(x_2) = IG(x_3)$ شد، یکی را به صورت تصادفی انتخاب می‌کنیم (مثلاً x_2). آخرین گره درخت هم به طور یکتا بدست می‌آید و درخت نهایی برای درس معماری کامپیوتر به شکل زیر در می‌آید:



حال با استفاده از این درخت می‌توان پیش‌بینی کرد که تمرین ششم معماری کامپیوتر تمديد می‌شود.
حالا همین کار را برای درس یادگیری ماشین انجام می‌دهیم:

$$H(y) = - \sum_i P(y = y_i) \log(P(y = y_i)) = -\frac{3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} \approx 0.292$$

$$\text{Threshold } x_1 \leq 5 : H(y|x_1) = -\frac{1}{5}(1 \log 1 + 0 \log 0) - \frac{4}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) \approx 0.24$$

$$\text{Threshold } x_1 \leq 6 : H(y|x_1) = -\frac{2}{5}(1 \log 1 + 0 \log 0) - \frac{3}{5}(\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3}) \approx 0.166$$

$$\text{Threshold } x_1 \leq 7 : H(y|x_1) = -\frac{2}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{3}{5}(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3}) \approx 0.286$$

$$\text{Best threshold for } x_1 \text{ is } 6 \implies IG(x_1) \approx 0.126$$

$$H(y|x_2) = -\frac{2}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{3}{5}(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3}) \approx 0.286$$

$$IG(x_2) \approx 0.006$$

$$\text{Threshold } x_3 \leq 0.4 : H(y|x_3) = -\frac{4}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{5}(1 \log 1 + 0 \log 0) \approx 0.24$$

$$\text{Threshold } x_3 \leq 0.5 : H(y|x_3) = -\frac{2}{5}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{3}{5}(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3}) \approx 0.286$$

$$\text{Best threshold for } x_3 \text{ is } 0.5 \implies IG(x_3) \approx 0.52$$

از آنجایی که $IG(x_1)$ از بقیه بیشتر است، پس متغیر x_1 در ریشه قرار می‌گیرد. حال ادامه درخت را می‌سازیم:

$$H(y) = - \sum_i P(y = y_i) \log(P(y = y_i)) = -\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} \approx 0.276$$

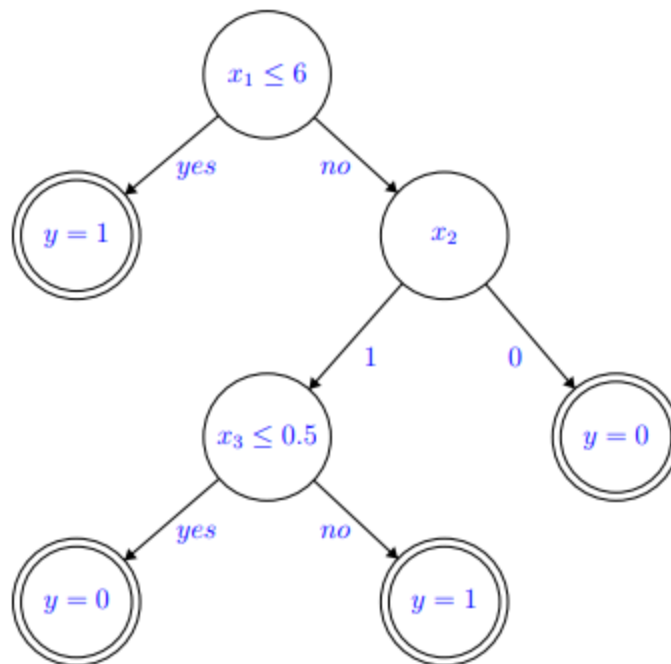
$$H(y|x_2) = -\frac{2}{3}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{3}(1 \log 1 + 0 \log 0) \approx 0.2$$

$$IG(x_2) \approx 0.76$$

$$\text{Threshold } x_3 \leq 0.5 : H(y|x_3) = -\frac{2}{3}(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) - \frac{1}{3}(1 \log 1 + 0 \log 0) \approx 0.2$$

$$IG(x_3) \approx 0.76$$

از آنجایی که $IG(x_1) = IG(x_2) = IG(x_3)$ شد، یکی را به صورت تصادفی انتخاب می‌کنیم (مثلاً x_1). آخرین گره درخت هم به طور یکتا بدست می‌آید و درخت نهایی برای درس یادگیری ماشین به شکل زیر در می‌آید:



حال با استفاده از این درخت می‌توان پیش‌بینی کرد که تمرین ششم یادگیری ماشین تمدید نمی‌شود. پس دانشجو باید هرچه سریع‌تر شروع به حل تمرین درس یادگیری ماشین کند.

شبکه‌های عصبی

1.

علت استفاده از لایه‌های کاملاً متصل در انتهای شبکه‌های عصبی پیچشی چیست؟

2.

منظور از همبستگی مکانی در تصویر چیست؟ توضیح دهید.

3.

آیا بیش از اندازه زیاد بودن نرخ آموزش، می‌تواند به واگرایی منجر شود و مقدار تابع هزینه را به بی‌نهایت میل دهد؟ توضیح دهید.

4.

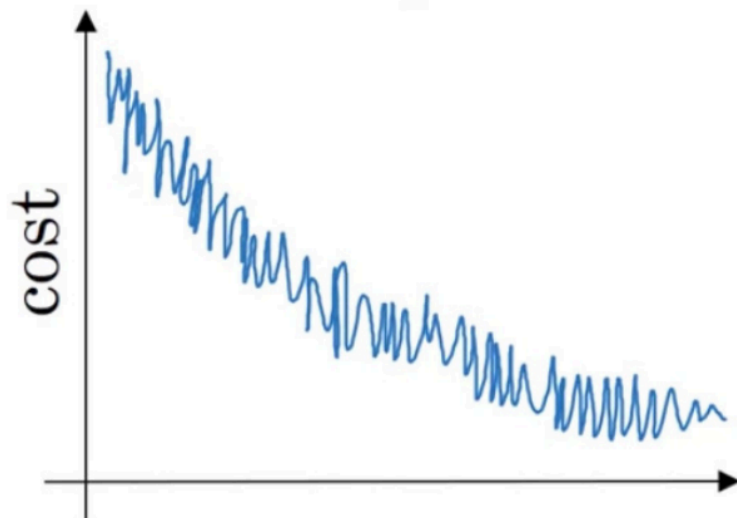
با تابع فعال‌ساز ReLU در درس آشنا شده‌اید. تابع فعال‌سازی دیگری به اسم LeakyReLU اینگونه تعریف می‌شود: $LeakyReLU(x) = \max(ax, x)$ که ضریب a در آن یک عدد کوچک مانند 0.01 می‌باشد.

الف) این تابع را رسم کنید.

ب) به نظر شما کدام یک از توابع ReLU و LeakyReLU کمتر با مشکل vanishing gradients مواجه می‌شود؟ دلایل خود را بیان کنید.

5.

فرض کنید نمودار هزینه-زمان شبکه شما در حین آموزش به شکل زیر باشد:



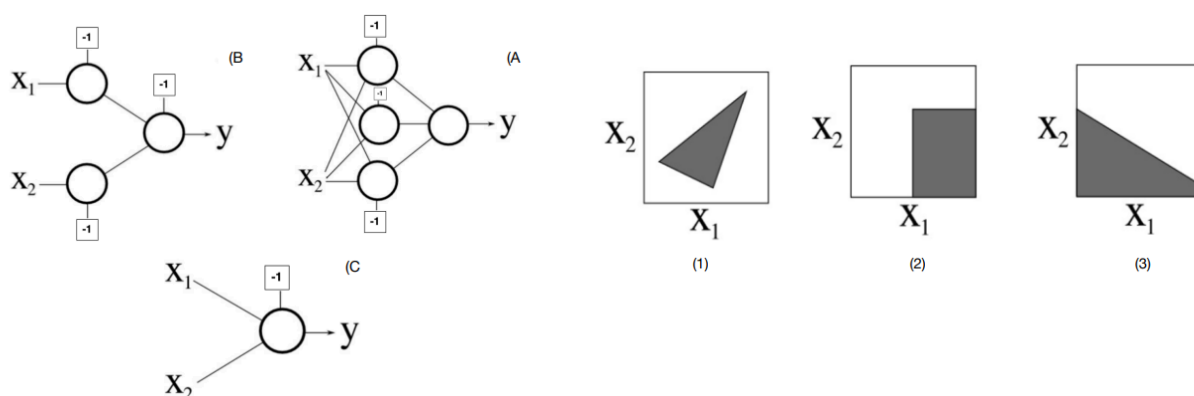
با فرض درست بودن نحوه آموزش، به نظر شما در آموزش از batch (بروزرسانی پارامترهای شبکه بعد از مشاهده همه داده‌ها) استفاده شده است یا از mini-batch (بروزرسانی پارامترهای شبکه بعد از مشاهده دسته‌ای از داده‌ها) استفاده شده است؟

6.

متین که به تازگی درس هوش مصنوعی را فراگرفته است در حال آموزش یک شبکه عصبی است. او که به نظر خودش بسیار در این کار مهارت دارد، تابع فعال‌ساز خودش را تعریف کرده است! این تابع برابر $f(x) = \text{floor}(e^x)$ می‌باشد. او ادعا می‌کند که این تابع غیرخطی است و شرایط لازم برای تابع فعال‌ساز بودن را دارد. همچنین او می‌خواهد از بهینه‌ساز Stochastic Gradient Descent برای آموزش مدل خود استفاده کند. با توجه به تابع فعال‌ساز و بهینه‌ساز استفاده شده، به نظر شما او موفق خواهد بود؟

7.

برای جداسازی نواحی هر تصویر، از کدام شبکه عصبی استفاده می‌کنید؟ دلایل خود را بیان کنید.



8.

هادی می‌خواهد یک شبکه عصبی آموزش دهد تا درصد احساسات موجود در یک جمله را تشخیص دهد. برای مثال یک داده به همراه لیبل آن در دیتاست او اینگونه می‌باشد:

```
{
  داده: "میدونستید یکشنبه امتحان پایان‌ترم هوش مصنوعی دارید!؟"
  %لیبل: شغف: 25% - تعجب: 50% - خشم: 15% - غم: 0% - اعتماد: 10% - ترس: 0
}
```

تابع فعال‌ساز softmax را توضیح دهید و بگویید به نظر شما هادی می‌تواند از این تابع در لایه آخر شبکه خودش استفاده کند؟