

به نام خدا



هوش مصنوعی

تمرین هفتم

استاد :

مهدی سمیعی

نویسنده :

محمد هومان کشوری

شماره دانشجویی :

۹۹۱۰۵۶۶۷

سوال ۱

.۱

در صورتی که یکی از ورودی‌های سافت مکس به بی نهایت میل کند، در این صورت به این دلیل که خروجی آن متشکل از تقسیم عبارات نمایی این ورودی‌ها است پس قطعاً انقدر مخرج بزرگ شده که تمامی احتمالات به ۰ میل می‌کند بجز احتمالاتی که بی نهایت در صورت آن است که به ۱ میل می‌کند.

.۲

در صورتی که یک پرسپترون خالی داشته باشیم به علت ذات غیر خطی xor نمی‌توانیم آنرا مدل کنیم اما اگر مثلاً بتوانیم چند لایه perceptron داشته باشیم این کار را می‌توان انجام داد.
برای and می‌توان تابع خطی زیر را در نظر گرفت :

$$y = w_1 * x_1 + w_2 * x_2 + b \xrightarrow{w_1=w_2=1, b=-1} y = x_1 + x_2 - 1$$

برای or می‌توان تابع زیر را در نظر گرفت

$$y = w_1 * x_1 + w_2 * x_2 + b \xrightarrow{w_1=w_2=2, b=-1} y = x_1 + x_2 - 1$$

سوال ۲

۱.

ابتدا باید ببینیم کدام احتمالات به چه صورت هستند.

$$\begin{aligned}
 P(w_1|Spam) &= \frac{5}{6}, P(w_1|Ham) = \frac{1}{4} \\
 P(w_2|Spam) &= \frac{4}{6}, P(w_2|Ham) = \frac{1}{4} \\
 P(w_3|Spam) &= \frac{1}{6}, P(w_3|Ham) = \frac{3}{4} \\
 P(w_4|Spam) &= \frac{2}{6}, P(w_4|Ham) = \frac{3}{4} \\
 P(w_5|Spam) &= \frac{6}{6}, P(w_5|Ham) = \frac{2}{4} \\
 P(Spam) &= \frac{6}{10} \\
 P(Ham) &= \frac{4}{10} \\
 P(Spam|w_1, w_2, w_3, w_4, w_5) &= \frac{P(Spam) * \prod_{i=1}^5 P(w_i|Spam)}{P(Spam) * \prod_{i=1}^5 P(w_i|Spam) + P(Ham) * \prod_{i=1}^5 P(w_i|Ham)} = \frac{\frac{6}{10} * \frac{5}{6} * \frac{4}{6} * \frac{1}{6} * \frac{2}{6}}{\frac{6}{10} * \frac{5}{6} * \frac{4}{6} * \frac{1}{6} * \frac{2}{6} + \frac{4}{10} * \frac{1}{4} * \frac{1}{4} * \frac{3}{4} * \frac{3}{4} * \frac{2}{4}} \\
 &= 0.732 \xrightarrow{>0.5} Spam
 \end{aligned}$$

۲.

$$P(Spam|w_1, w_2, w'_3, w'_4, w'_5) = \frac{\frac{6}{10} * \frac{5}{6} * \frac{4}{6} * \frac{5}{6} * \frac{4}{6} * \frac{0}{6}}{X} = 0 \xrightarrow{<0.5} Ham$$

عملا داده آخر کار کلی را خراب کرده است چرا که به دلیل عدم smoothing یک ۰ در کل عبارات ضرب شده که اصلا مطلوب نیست.

۳.

در این حالت فرض می‌کنیم یک عدد از تمامی outcome ها حداقل دیده‌ایم.

$$P(Spam|w_1, w_2, w'_3, w'_4, w'_5) = \frac{\frac{11}{20} * \frac{6}{11} * \frac{5}{11} * \frac{9}{11} * \frac{8}{11} * \frac{4}{11}}{\frac{11}{20} * \frac{6}{11} * \frac{5}{11} * \frac{9}{11} * \frac{8}{11} * \frac{4}{11} + \frac{9}{20} * \frac{2}{9} * \frac{2}{9} * \frac{5}{9} * \frac{5}{9} * \frac{6}{9}} = 0.865 \xrightarrow{>.5} Spam$$

سوال ۳

۱.

$$\begin{aligned}\frac{d_{Cost}}{d_{h_{\theta}(x)}} &= -2(y - h_{\theta}(x)), \\ \frac{d(\sigma(z))}{dz} &= \frac{e^{-z}}{(1+e^{-z})^2} \\ \frac{d(z)}{d(w_i)} &= x_i \\ \frac{d_{Cost}}{d_{w_i}} &= -2(y - \sigma(z)) * \left(\frac{e^{-z}}{(1+e^{-z})^2}\right) * x_i\end{aligned}$$

حال در هر مرحله باید w_i ها را مطابق فرمول زیر محاسبه کنیم.

$$w_i = w_i + \eta * 2(y - \sigma(z(x))) * \left(\frac{e^{-z(x)}}{(1+e^{-z(x)})^2}\right) * x_i$$

۲.

ابتدا باید شرط convex بودن تابع بررسی شود که مطمئن باشیم به یک جواب بهینه می‌رسیم. که تابع اول این ویژگی را دارد. دومین علت آن است که در مقایسه با باقی function loss ها، regression logistic عملاً بهتر عمل می‌کند چرا که اثبات می‌شود که یک MLE است که یعنی بالاترین استاندارد برای تخمین و بایاس را دارد که یعنی با تقریب خوبی بهترین گزینه ما است.

سوال ۴

۱.

مفهوم $overfitting$ عملاً به این معنا است که انقدر نمودار نهایی را در هنگام رگرسیون به داده‌ها نزدیک کنیم که نمودار کلی ما برای تخمین داده‌های بعدی اشتباه باشد. مفهوم $underfitting$ نیز این است که آنقدر نمودار نهایی رگرسیون را ساده فرض کنیم که تخمین غلطی از داده‌ها به ما بدهد. در شکل مشاهده شده می‌توان دید پیچیدگی زیاد مدل منجر به $overfitting$ شده و هرچند ارور $Set Train$ کم شده است اما ارور $vaildation$ زیاد شده است که یعنی صرفاً داریم نمودار دقیق‌تر برای $Set Training$ می‌دهیم اما لزوماً پیش‌بینی درستی نداریم. در سمت چپ نمودار یعنی با مدل خیلی ساده نیز همانطور که انتظار می‌رود $underfitting$ داریم چرا که انگار یک نمودار خیلی ساده بدون در نظر گرفتن رفتارهای مدل داریم به آن نسبت می‌دهیم.

۲.

نمودار دوم چرا که اگر خیلی مدل پیچیده باشد دچار $overfitting$ شده و اگر نیز خیلی ساده باشد دچار $underfitting$ می‌شود که باید از هر دو اجتناب کنیم و یک حد خوب برای تخمین نمودار رگرسیون بدست آوریم.

سوال ۵

۱. بله چرا که هر چه داده‌های ما بیشتر باشند احتمالاً تخمین دقیق‌تری از تست خواهیم داشت.
۲. همواره نه چرا که اگر بیش از حد کاهش دهیم می‌تواند به overfitting منجر شود.
۳. همواره نه چرا که اگر خیلی بزرگ باشد می‌تواند به underfitting منجر شود.
۴. احتمالاً بهتر می‌کند اما نمی‌توان گفت به صورت خطی بهتر می‌کند.