به نام خدا
تکلیف سوم یادگیری ماشین
نیمسال تحصیلی ۰۱-۰۰
موعد تحویل: ۳۰ آذر ساعت ۲۳:۵۹

---

۱. ثابت کنید تابع softmax نسبت به اضافه شدن مقدار ثابت به ورودی، حساس نیست. به عبارت دیگر تساوی زیر برقرار است: ($x + c$ به معنای افزودن مقدار ثابت c به تمام ابعاد x می‌باشد.)

$$softmax(x) = softmax(x + c)$$

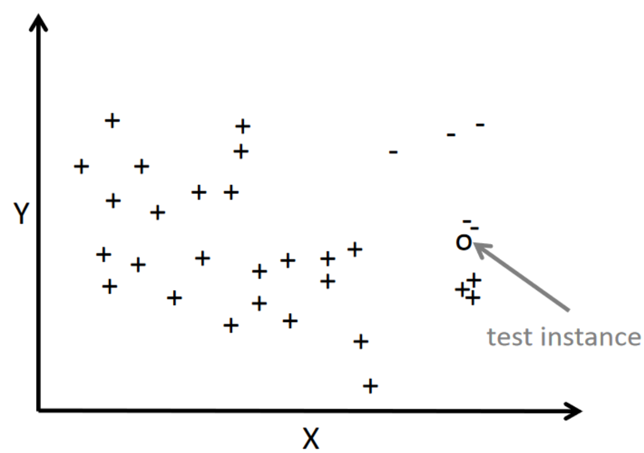
برای تابع softmax داریم:

$$softmax(x)_i = \frac{e^{x_i}}{\sum_j \quad e^{x_j}}$$

پاسخ: برای تمام ابعاد با شماره‌ی i که $1 \le i \le dim(x)$ داریم:

$$\left(softmax(x+c)\right)_i = \frac{exp(x_i + c)}{\sum_{j=1}^{dim(x)} exp(x_j + c)} = \frac{exp(c)exp(x_i)}{exp(c)\sum_{j=1}^{dim(x)} exp(x_j)} = \frac{exp(x_i)}{\sum_{j=1}^{dim(x)} exp(x_j)} = \left(softmax(x)\right)_i$$

۲. فرض کنید قصد داریم یک مدل K نزدیکترین همسایگی را بر روی داده‌های موجود در شکل زیر اجرا کنیم. شکل شامل داده‌هایی از دسته‌ی (+)، دسته‌ی (-) و یک داده به شکل (o) می‌باشد که همان داده تست ما خواهد بود. به هر یک از سوالات زیر با ذکر دلیل مناسب، پاسخ دهید.



الف) اگر در این مدل k برابر با ۱ در نظر گرفته شود، داده‌ی تست به کدام یک از دو کلاس (+) یا (-) تعلق خواهد یافت؟

پاسخ: KNN یک داده‌ی تست را به دسته‌ای با نزدیکترین k همسایه‌ی موجود در اطراف آن داده‌ی تست، تعلق می‌دهد. پس در این حالت داده‌ی تست به دسته‌ی (-) تعلق می‌گیرد چون نزدیکترین همسایه‌ی آن از نوع (-) می‌باشد.

ب) اگر در این مدل k برابر با ۳ در نظر گرفته شود، دادهی تست به کدام یک از دو کلاس (+) یا (-) تعلق خواهد یافت؟

پاسخ: KNN یک دادهی تست را به دستهای با نزدیکترین k همسایهی موجود در اطراف آن دادهی تست، تعلق میدهد. پس در این حالت دادهی تست به دستهی (-) تعلق میگیرد چون از ۳ تا نزدیکترین همسایهی نقطهی تست، دو نقطه از نوع (-) و 1 نقطه از نوع (+) میباشد.

ج) اگر در این مدل k را بزرگتر ۱۰ در نظر بگیریم، دادهی تست به کدام یک از دو کلاس (+) یا (-) تعلق خواهد یافت؟

پاسخ: در تصویر تنها تعداد ۵ داده از نوع (-) میباشند، پس اگر k>10، در هر حالتی نقطهی تست به دستهی (+) تعلق خواهد یافت.

۳. جدول زیر یک مجموعه آموزشی شامل ۸ نمونه میباشد. در این جدول چهار ستون smelly، height، legs و color ویژگیهای هر نمونه (features) هستند. ستون Species نیز ستون هدف (target) میباشد که دارای دو دستهی M و H است. با استفاده از طبقهبند Naïve Bayes، محاسبه کنید که نمونه دادهی زیر به کدام یک از دو دستهی M یا H تعلق خواهد گرفت.

(color = green, legs = 2, height = tall, smelly = no)

ID	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

پاسخ:

لازم است هر دو حالتی که Species برابر با M یا برابر با H باشد را بررسی کنیم تا پیشبینی نهایی به دست آید:

$X = (color = green, legs = 2, height = tall, smelly = no)$

حالت اول: Species=M

$$P(X|Species = M) = P(color = green|M) \times P(legs = 2|M) \times P(height = tall|M) \times P(smelly = no|M) = \frac{2}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{2}{256}$$

حالت دوم: Species=H

$$P(X|Species = H) = P(color = green|H) \times P(legs = 2|H) \times P(height = tall|H) \times P(smelly = no|H) = \frac{1}{4} \times \frac{4}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{24}{256}$$

پس:

$$P(X|Species = M)P(Species = M) = \frac{2}{256} \times \frac{4}{8} = 0.0039$$

$$P(X|Species = H)P(Species = H) = \frac{24}{256} \times \frac{4}{8} = 0.0468$$

همانطور که مشاهده می‌شود:

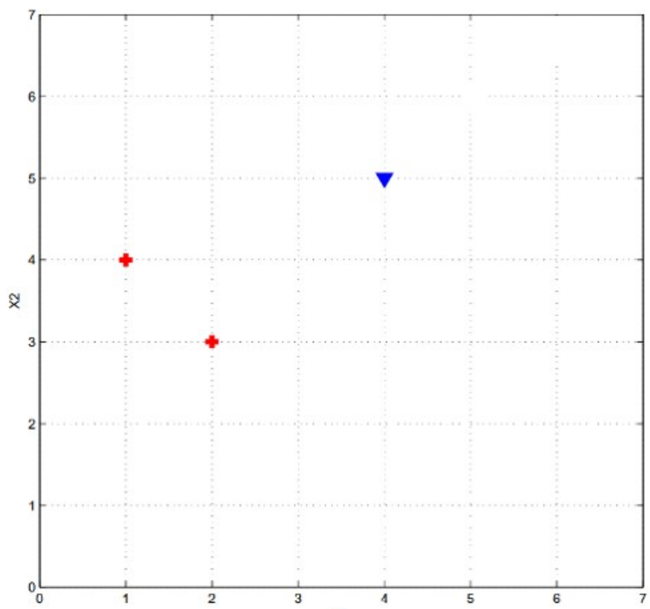
$$P(X|Species = M)P(Species = M) < P(X|Species = H)P(Species = H)$$

بنابراین:

$$P(Species = M|X) < P(Species = H|X)$$

پس این نمونه داده، متعلق به دسته‌ی H می‌باشد.

۴. می‌خواهیم یک طبقه‌بند ماشین بردار پشتیبان را روی داده‌های زیر آموزشی زیر آموزش دهیم. در این شکل، ۲ داده با مقدار ۱- (مثبت های قرمز) و ۱ داده با مقدار ۱+ (مثلث آبی) داریم. (حل سوال به صورت تحلیلی است.)



الف) معادله خط تصمیم چه خواهد بود؟ (مقادیر w، b و مارجین یا m را به دست آورید)

$$L(w,b,\alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i \left(y_i (x_i^T w + b) - 1\right) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i y_i x_i^T w - b\left(\sum_{i=1}^{n} \alpha_i y_i\right) + \sum_{i=1}^{n} \alpha_i$$

① $\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i \Rightarrow w^* = \sum_{i=1}^{n} \alpha_i y_i x_i$ $\xrightarrow{\text{جای‌گذاری}}$ $w^* = -\alpha_1 \begin{bmatrix}1\\4\end{bmatrix} - \alpha_2 \begin{bmatrix}2\\3\end{bmatrix} + \alpha_3 \begin{bmatrix}4\\5\end{bmatrix}$

② $\frac{\partial L}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = \emptyset$ $\qquad -\alpha_1 - \alpha_2 + \alpha_3 = 0$

دوگان

$$\max_{\alpha_1,\alpha_2,\dots,\alpha_n} L(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)} - b \cancel{\sum_{i=1}^{n} \alpha_i y^{(i)}}^{\emptyset}$$

$$\sum \alpha_i y_i = \emptyset \;,\; \alpha_i \geq 0 \;,\; i = 1,2,\dots,n$$

$$\max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{3} \alpha_i$$

$$x_1^T x_1 = 17 \;,\; x_1^T x_2 = 14 \;,\; x_1^T x_3 = 24 \;,\; x_2^T x_2 = 13 \;,\; x_2^T x_3 = 23 \;,\; x_3^T x_3 = 41$$

$$\begin{cases} \max_{\alpha_1,\alpha_2,\alpha_3} -\frac{1}{2}\left[17\alpha_1^2 + 28\alpha_1\alpha_2 - 48\alpha_1\alpha_3 + 13\alpha_2^2 - 46\alpha_2\alpha_3 + 41\alpha_3^2\right] + \alpha_1 + \alpha_2 + \alpha_3 \\ -\alpha_1 - \alpha_2 + \alpha_3 = 0 \Rightarrow \alpha_3 = \alpha_1 + \alpha_2 \end{cases}$$

$$y = \max_{\alpha_1,\alpha_2} -\frac{1}{2}\left[10\alpha_1^2 + 16\alpha_1\alpha_2 + 8\alpha_2^2\right] + 2\alpha_1 + 2\alpha_2$$

$$\begin{cases} \frac{\partial y}{\partial \alpha_1} = -10\alpha_1 - 8\alpha_2 + 2 = 0 \\ \frac{\partial y}{\partial \alpha_2} = -8\alpha_1 - 8\alpha_2 + 2 = 0 \end{cases} \Rightarrow \begin{matrix} \alpha_1 = 0 \\ \alpha_2 = 0.25 \\ \alpha_3 = 0.25 \end{matrix}$$

$$w^* = \sum_{i=1}^{3} \alpha_i y_i x^{(i)} = -0.25\begin{bmatrix}2\\3\end{bmatrix} + 0.25\begin{bmatrix}4\\5\end{bmatrix} = \begin{bmatrix}0.5\\0.5\end{bmatrix}$$

$$b^* \Rightarrow y_2(x_2^T w + b^*) = 1 \qquad b^* = -1 - \begin{bmatrix}2\\3\end{bmatrix}^T \begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\end{bmatrix} = -3.5 \qquad m = \frac{2}{\|w\|} = 2\sqrt{2}$$

ب) نقاط بردار پشتیبان را روی عکس مشخص کنید و خط تصمیم را رسم کنید.

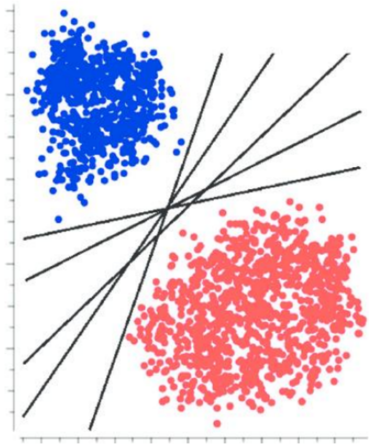
۵. در مسئله regularized logistic regression زیر، فرض کنید j میتواند یکی از سه مقدار 0،1،2 باشد(به عبارت دیگر بردارθ یک بردار با ابعاد یک در سه است) با توجه به داده های آموزشی زیر، توضیح دهید بعد از منتظم سازی با مقادیر بزرگ λ، به ازای هر پارامتر، خطای آموزش چه تغییری می‌کند(به عبارت دیگر با منتظم سازی $\theta 0$ میزان خطا چه تغییری میکند، و سپس به ترتیب منتظم سازی $\theta 1$ و $\theta 2$). درباره تغییرات هر مورد توضیح دهید.

پاسخ:

الف) منظم سازی روی $\theta 0$: در این حالت مقدار bias به سمت صفر میرود و در نهایت خط مرزی از مبدا مختصات عبور خواهد کرد. از آنجایی که میتوان خطی کشید که از مبدا عبور کند و داده ها را به خوبی تفکیک کند، مقدار خطای آموزشی تغییری نمیکند.

ب) منظم سازی روی $\theta 1$: در این حالت منظم سازی روی w1 انجام میشود که باعث میل کردن w1 به سمت صفر میشود. در نتیجه خط مرزی افقی خواهد شد که خطی افقی مناسبی برای تفکیک داده وجود ندارد. در نتیجه خطای آموزشی زیاد خواهد شد.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^i(\log\left(h_\theta(x^i)\right) - (1-y^i)(\log\left(1-h_\theta(x^i)\right)\right] + \lambda\theta_j^2$$

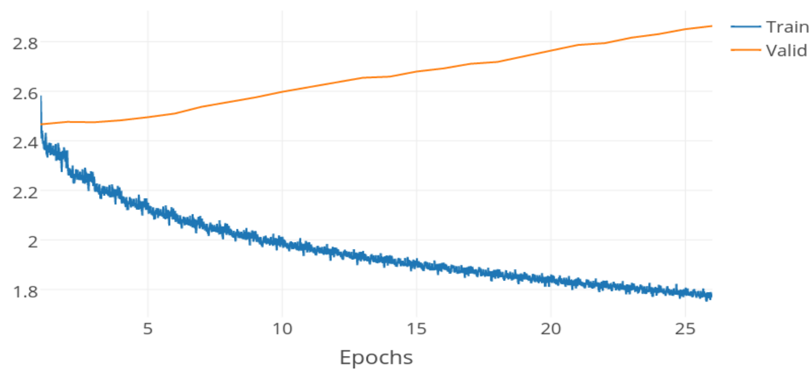


۶. فرض کنید یک مدل رگرسیون لاجستیک برای تشخیص بیماران سرطان طراحی کرده اید و پس از آموزش شبکه، منحنی های آموزش زیر مشاهده شده است. ابتدا توضیح دهید مدل از چه مشکلی رنج می برد و سپس بگویید کدام یک از موارد زیر می تواند به بهبود مدل کمک کند. در هر مورد، توضیح کوتاهی ارائه دهید و مشخص کنید مقدار bias و variance بعد از اجرای هر کدام از این موارد چه تغییری خواهد کرد.

الف) اضافه کردن ویژگی های جدید.

ب) بزرگتر کردن مجموعه آموزشی.

ج) بزرگتر کردن پارامتر منتظم‌سازی.



پاسخ: از آنجایی که ارور شبکه روی دیتای train بسیار کم و برای دیتای test بسیار زیاد است، مدل variance زیادی روی دیتا آموزشی پیدا کرده است و توان تشخیص دیتای test را ندارد.(overfit شده است)

الف) این روش کمک کننده نیست. با اضافه کردن feature های جدید، مدل پیچیده تر میشود و variacne بیشتری روی دیتای آموزشی پیدا میکند و مشکل شبکه بدتر میشود. پس نمیتوان از این روش استفاده کرد. در نهایت مدل دارای واریانس بیشتری و bias کمتری خواهد بود.

ب) این روش میتواند کمک کننده باشد. بیشتر کردن داده های آموزشی میتواند واریانس داده ها را زیاد تر و variance روی داده های آموزشی را کمتر میکند که باعث حل مشکل overfit میشود. در نهایت مدل دارای واریانس کمتر و bias بیشتری خواهد بود.

ج) این روش میتواند کمک کننده باشد. با بزرگتر کردن λ، میزان خطا روی بزرگ بودن وزن ها بیشتر میشود و باعث کمک به کمتر کردن variance شبکه و در نتیجه حل مشکل overfit شود . در نهایت مدل دارای واریانس کمتر و bias بیشتری خواهد بود.