تکلیف سری سوم درس یادگیری ماشین

سارا برادران (شماره دانشجویی: ۹۶۲۴۱۹۳)

Question 1)

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

Question 2)

الف) اگر فاصله اقلیدسی داده تست از سایر داده ها را در نظر بگیریم و k=1 باشد آن گاه تنها نزدیک ترین داده به داده تست مد نظر قرار داده می شود و چون نزدیک ترین داده به داده تست (برحسب فاصله اقلیدسی) متعلق به کلاس منفی است لذا داده تست نیز جزوه داده های منفی (–) طبقه بندی می شود.

ب) اگر فاصله اقلیدسی داده تست از سایر داده ها را در نظر بگیریم و k=2 باشد آن گاه تنها نزدیک ترین دو داده به داده تست مد نظر قرار داده می شود و چون نزدیک ترین دو داده به داده تست (برحسب فاصله اقلیدسی) هر دو متعلق به کلاس منفی هستند لذا داده تست نیز جزوه داده های منفی (–) طبقه بندی می شود.

ج) اگر فاصله اقلیدسی داده تست از سایر داده ها را در نظر بگیریم و k>10 باشد آن گاه بیش از ۱۰ مورد از نزدیک ترین داده ها به داده تست مد نظر قرار داده می شود و اگر بررسی کنیم مشخص است که در مجموع تنها ۵ مورد داده متعلق به کلاس منفی بوده و سایر داده ها متعلق به کلاس مثبت می باشد لذا اگر بیش از ۱۰ مورد از نزدیک ترین داده ها به داده تست را انتخاب نماییم به قطع تعداد داده های کلاس مثبت در آن حداقل ۶ مورد می باشد در حالی که حداکثر ۵ داده از کلاس منفی داریم و لذا داده تست نیز جزو داده های مثبت (+) طبقه بندی می شود.

Question 3)

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

P (species = M | X) = $\frac{\text{P(species=M and X)}}{\text{P(x)}} = \frac{\text{P(species=M)}*\text{P(X|species=M)}}{\text{P(x)}}$

P (X | species = M) = P (color = green | species = M) $*$ P (legs = 2 | species = M) $*$ P (height = tall | species = M) $*$ P (smelly = No | species = M)

P (color = green | species = M) = $\frac{2}{4}$

P (legs = 2 | species = M) = $\frac{1}{4}$

P (height = tall | species = M) = $\frac{1}{4}$

P (smelly = No | species = M) = $\frac{1}{4}$

P (species = M) = $\frac{4}{8}$

$$P\ (species = M \mid X) = \frac{\frac{4}{8}*(\frac{2}{4}*\frac{1}{4}*\frac{1}{4}*\frac{1}{4})}{\frac{1}{8}} = \frac{1}{32}$$

$$P\ (species = H \mid X) = \frac{P(species{=}H \text{ and } X)}{P(x)} = \frac{P(species{=}H)*P(X|species{=}H)}{P(x)}$$

$P\ (X \mid species = H) = P\ (color = green \mid species = H) * P\ (legs = 2 \mid species = H) * P\ (height = tall \mid species = H) * P\ (smelly = No \mid species = H)$

$$P\ (color = green \mid species = H) = \frac{1}{4}$$

$$P\ (legs = 2 \mid species = H) = \frac{4}{4}$$

$$P\ (height = tall \mid species = H) = \frac{2}{4}$$

$$P\ (smelly = No \mid species = H) = \frac{3}{4}$$

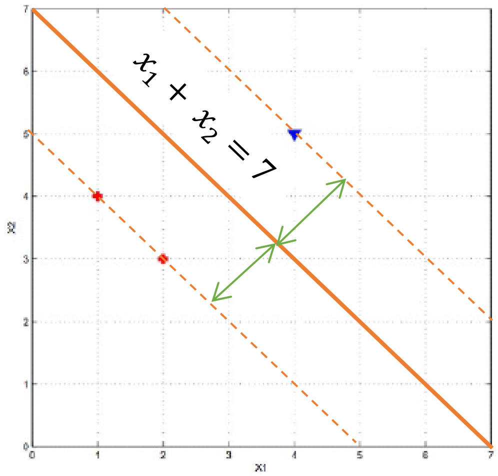
$$P\ (species = H) = \frac{4}{8}$$

$$P\ (species = H \mid X) = \frac{\frac{4}{8}*(\frac{1}{4}*\frac{4}{4}*\frac{2}{4}*\frac{3}{4})}{\frac{1}{8}} = \frac{12}{32}$$

$$P\ (species = H\ ,\ X) > P\ (species = M\ ,\ X) \rightarrow P\ (species = H \mid X) > P\ (species = M \mid X)$$

نمونه ی داده شده به دسته H تعلق خواهد گرفت.

Question 4)



بردار های پشتیبان رسم شده (طول فواصل سبز رنگ نشان داده شده برابر $\sqrt{2}$ می باشد)

X^1 (4,5) y^1 = +1

X^2 (2,3) y^2 = -1

X^3 (1,4) y^3 = -1

$$L(w.b.\alpha) = \frac{1}{2}\ ||w||_2^2 - \sum_{i=1}^{3} \alpha_i(y_i(x_i^T w + b) - 1)$$

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

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{3} \alpha_i y_i = \alpha_1 - \alpha_2 - \alpha_3 = 0$$

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

$x_1^T x_1 = 41 \qquad x_1^T x_3 = 24 \qquad x_2^T x_3 = 14$

$x_1^T x_2 = 23 \qquad x_2^T x_2 = 13 \qquad x_3^T x_3 = 17$

Max $\alpha_1.\alpha_2.\alpha_3 \quad -\frac{1}{2}[41\alpha_1^2 - 46\alpha_1\alpha_2 - 48\alpha_1\alpha_3 + 13\alpha_2^2 + 28\alpha_2\alpha_3 + 17\alpha_3^2] + \alpha_1 + \alpha_2 + \alpha_3$

$\alpha_1 - \alpha_2 - \alpha_3 = 0$

$\alpha_1 - \alpha_2 = \alpha_3$

J = Max $\alpha_1.\alpha_2 \quad -\frac{1}{2}[41\alpha_1^2 - 46\alpha_1\alpha_2 - 48\alpha_1(\alpha_1 - \alpha_2) + 13\alpha_2^2 + 28\alpha_2(\alpha_1 - \alpha_2) + 17(\alpha_1 - \alpha_2)^2] + \alpha_1 + \alpha_2 + \alpha_1 - \alpha_2 = -\frac{1}{2}[41\alpha_1^2 - 46\alpha_1\alpha_2 - 48\alpha_1^2 + 48\alpha_1\alpha_2 + 13\alpha_2^2 + 28\alpha_1\alpha_2 - 28\alpha_2^2 + 17\alpha_1^2 + 17\alpha_2^2 - 34\alpha_1\alpha_2] + 2\alpha_1 = -\frac{1}{2}[10\alpha_1^2 + 2\alpha_2^2 - 4\alpha_1\alpha_2] + 2\alpha_1$

$$\frac{\partial L}{\partial \alpha_1} = -10\alpha_1 + 2\alpha_2 + 2 = 0$$

$$\frac{\partial L}{\partial \alpha_2} = -2\alpha_2 + 2\alpha_1 = 0 \rightarrow \alpha_2 = \alpha_1$$

$$-10\alpha_1 + 2\alpha_1 + 2 = 0 \rightarrow -8\alpha_1 = -2$$

$$\rightarrow \alpha_1 = \frac{1}{4}\ .\alpha_2 = \frac{1}{4}.\alpha_3 = 0$$

$$w^* = \sum_{i=1}^{3} \alpha_i y_i x_i = \frac{1}{4}\begin{bmatrix}4\\5\end{bmatrix} - \frac{1}{4}\begin{bmatrix}2\\3\end{bmatrix} - 0\begin{bmatrix}1\\4\end{bmatrix} = \begin{bmatrix}0.5\\0.5\end{bmatrix}$$

$$m = \frac{2}{||w||} = 2\sqrt{2}$$

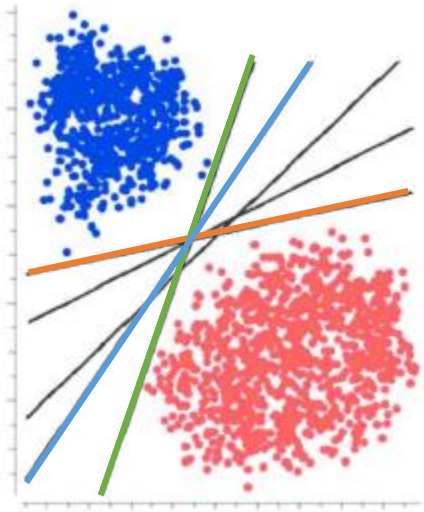
$$b^* = y_2(x_2^T w + b^*) = 1 \rightarrow -\left([2\ 3]\begin{bmatrix}0.5\\0.5\end{bmatrix} + b^*\right) = 1 \rightarrow b^* = -\frac{7}{2}$$

$$wx + b = 0 \to \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} [x_1\ x_2] - \frac{7}{2} = 0 \to \frac{1}{2}x_1 + \frac{1}{2}x_2 = \frac{7}{2} \to x_1 +\ x_2 = 7$$

Question 5)

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

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



۱- منظم سازی بر روی θ_2 : هنگامی که منظم سازی بر روی θ_2 صورت گیرد مرز تصمیم وابستگی کمتری نسبت به x_2 پیدا می کند و لذا عمودی می شود. (مشابه خط سبز رنگ در تصویر)

$\theta_0 + \theta_1 x_1 = 0 \to x_1 = -\frac{\theta_0}{\theta_1} \to$ مرز تصمیم عمودی

در این حالت برای مقادیر بزرگ λ خطای آموزش افزایش می یابد چراکه هیچ خط عمودی به عنوان مرز جدا کننده نمی توان یافت که داده ها را به خوبی از یکدیگر تفکیک نماید.

۲- منظم سازی بر روی θ_1 : هنگامی که منظم سازی بر روی θ_1 صورت گیرد خطای آموزش افزایش یافته و مرز تصمیم وابستگی کمتری به x_1 پیدا می کند و لذا افقی می شود. (مشابه خط نارنجی رنگ در تصویر) در چنین حالتی داده های آموزشی توسط یک خط افقی از هم جدا شده و خطای بیشتری در دسته بندی رخ می دهد. البته این در حالتی است که خط بهینه خطی مورب باشد و در صورتی که خط بهینه افقی باشد می توان ادعا کرد با منظم سازی بر روی θ_1 خطای آموزشی تغییری نمی کند.

۳- منظم سازی بر روی θ_0 : در این حالت خطای آموزش تغییری نمی کند. به سبب اینکه بایاس به سمت صفر میل می کند مرز تصمیم به سمت مبدا مختصات رفته و خط بهینه برای جدا سازی داده ها ممکن است از مبدا مختصات عبور کند اما برای داده های آموزش در شکل داده شده می توان خطی با خطای صفر در نظر گرفت که از مبدا مختصات نیز عبور می کند.(مشابه خط آبی رنگ در تصویر)

Question 6)

خطای حاصل از به کارگیری مدل روی داده‌های آموزشی، بایاس مدل گفته شده و از طرفی خطای مدل روی داده‌های آزمایشی، واریانس مدل در نظر گرفته می‌شود. با توجه به تصویر نمودار های رسم شده خطای مدل بر روی داده های آموزشی کم و خطای مدل بر روی داده های تست زیاد بوده است لذا می توان گفت بایاس کم و واریانس زیاد است (مدل high variance است.) به عبارت دیگر واریانس بزرگ زمانی رخ می دهد که مدل نسبتا پیچیده شده و بیش برازش (overfitting) بر روی داده های آموزشی رخ دهد یکی از دلایل رخداد این مشکل عدم تناسب در حجم داده های تست و آموزشی است (مثلا زمانی که حجم داده آموزشی کم و حجم داده تست زیاد باشد). قسمت های ب و ج می توانند منجر به کاهش واریانس شده و مدل را بهبود بخشند.

الف) اینکار می تواند منجر به افزایش واریانس و کاهش بایاس شود و مسئله بیش برازش را تشدید می کند.

ب) اینکار می تواند منجر به کاهش واریانس و افزایش بایاس شود و برای بهبود مدل مناسب است. چراکه یکی از دلایل رخداد high variance و یا بیش برازش عدم تناسب میان حجم داده آموزشی و داده تست است و چنانچه حجم داده آموزشی افزایش یابد مدل قابلیت تعمیم می یابد و مسئله overfitting حل می گردد.

ج) اینکار می تواند منجر به کاهش واریانس و افزایش بایاس شود و برای بهبود مدل مناسب است. (واضح است که تنظیم ضریب regularization می تواند منجر به بهبود مدل و کاهش خطا بر روی داده های تست گردد و به عبارت دیگر مسئله overfitting را برطرف سازد.)