تکلیف سری دوم درس یادگیری ماشین

سارا برادران (شماره دانشجویی: ۹۶۲۴۱۹۳)

سوال ۱)

الف)

$$L(\theta) = \prod_{i=1}^{n} p^{Xi}(1-p)^{1-Xi}$$

$$l(p) = \sum_{i=1}^{n} \log\left(p^{Xi}(1-p)^{1-Xi}\right) = \sum_{i=1}^{n} \log\left(p^{Xi}\right) + \log((1-p)^{1-Xi})$$

$$= \sum_{i=1}^{n} xi * \log(p) + (1-xi) * \log(1-p)$$

$$= \log(p) \sum_{i=1}^{n} xi + \log(1-p) \sum_{i=1}^{n}(1-xi) = \log(p) * \left(\sum_{i=1}^{n} xi\right)$$

$$+ \log(1-p)\left(n - \sum_{i=1}^{n} xi\right)$$

$$\frac{\partial l(p)}{\partial p} = \frac{\sum_{i=1}^{n} xi}{p} - \frac{(n - \sum_{i=1}^{n} xi)}{1-p} = 0 \rightarrow \frac{(1-p)(\sum_{i=1}^{n} xi) - p(n - \sum_{i=1}^{n} xi)}{p(1-p)}$$

$$= \frac{(\sum_{i=1}^{n} xi) - p(\sum_{i=1}^{n} xi) - pn + p(\sum_{i=1}^{n} xi)}{p(1-p)}$$

$$= \frac{(\sum_{i=1}^{n} xi) - p\left((\sum_{i=1}^{n} xi) + n - (\sum_{i=1}^{n} xi)\right)}{p(1-p)} = \frac{(\sum_{i=1}^{n} xi) - pn}{p(1-p)} = 0$$

$$\rightarrow \left(\sum_{i=1}^{n} xi\right) - pn = 0 \rightarrow p = \frac{1}{n}\left(\sum_{i=1}^{n} xi\right)$$

ب)

$$f(x|\alpha.\beta) \frac{1}{B(\alpha.\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$l(p) + \log\left(f(p|\alpha.\beta)\right)$$

$$= \log(p) \sum_{i=1}^{n} xi + \log(1-p) \sum_{i=1}^{n}(1-xi) = \log(p) * \left(\sum_{i=1}^{n} xi\right)$$

$$+ \log(1-p)\left(n - \sum_{i=1}^{n} xi\right) + \log\left(\frac{1}{B(\alpha.\beta)}(p)^{\alpha-1}(1-p)^{\beta-1}\right) = \log(p) * \left(\sum_{i=1}^{n} xi\right)$$

$$+ \log(1-p)\left(n - \sum_{i=1}^{n} xi\right) + \log\left(\frac{1}{B(\alpha.\beta)}\right) + \log((p)^{\alpha-1}) + \log\left((1-p)^{\beta-1}\right)$$

$$= \log(p) * \left(\sum_{i=1}^{n} xi\right) + \log(1-p)\left(n - \sum_{i=1}^{n} xi\right) + \log\left(\frac{1}{B(\alpha.\beta)}\right) + (\alpha-1)\log(p)$$

$$+ (\beta-1)\log(1-p)$$

$$= \log(p)\left(\alpha - 1 + \left(\sum_{i=1}^{n} xi\right)\right) + \log(1-p)\left(\beta - 1 + n - \sum_{i=1}^{n} xi\right) + \log\left(\frac{1}{B(\alpha.\beta)}\right)$$

می بایست بیشینه مقدار تابع فوق را بدست آوریم لذا نسبت به p مشتق گرفته و برابر صفر قرار می دهیم تا بررسی شود به ازای کدام p مقدار $l(p)+\log(\,f(p|\alpha.\beta)\,)$ بیشینه می گردد.

$$\frac{\partial[l(p)+\log(\,f(p|\alpha.\beta)\,)]}{\partial p}=0\rightarrow\ \frac{\left(\alpha-1+(\sum_{i=1}^{n}xi)\right)}{p}-\frac{(\beta-1+n-\sum_{i=1}^{n}xi)}{1-p}=0$$

$$\rightarrow\ \frac{(1-p)\left(\alpha-1+(\sum_{i=1}^{n}xi)\right)-\ p(\beta-1+n-\sum_{i=1}^{n}xi)}{p(1-p)}=\ 0$$

$$\rightarrow\left(\alpha-1+\left(\sum_{i=1}^{n}xi\right)\right)-p\left(\alpha-1+\left(\sum_{i=1}^{n}xi\right)\right)-p\left(\beta-1+n-\sum_{i=1}^{n}xi\right)=0$$

$$\rightarrow\left(\alpha-1+\left(\sum_{i=1}^{n}xi\right)\right)-p\left(\left(\alpha-1+\left(\sum_{i=1}^{n}xi\right)\right)+\left(\beta-1+n-\sum_{i=1}^{n}xi\right)\right)=0$$

$$\rightarrow p=\frac{\left(\alpha-1+(\sum_{i=1}^{n}xi)\right)}{\left(\alpha-1+\left(\sum_{i=1}^{n}xi\right)\right)+\left(\beta-1+n-\sum_{i=1}^{n}xi\right)}=\frac{\left(\alpha-1+(\sum_{i=1}^{n}xi)\right)}{\alpha+\beta+n-2}$$

سوال ۲)

الف)

$$\hat{y}^{(1)}=\ w[0]+w[1]*41+w[2]*138=\ -59.5+\ -0.15*41+0.60*138=17.15$$

$$\hat{y}^{(2)}=\ w[0]+w[1]*42+w[2]*153=\ -59.5+\ -0.15*42+0.60*153=26.00$$

$$\hat{y}^{(3)}=\ w[0]+w[1]*37+w[2]*151=\ -59.5+\ -0.15*37+0.60*151=25.55$$

$$\hat{y}^{(4)}=\ w[0]+w[1]*46+w[2]*133=\ -59.5+\ -0.15*46+0.60*133=13.40$$

$$MSE=\frac{1}{2m}\sum_{i=1}^{m}(\hat{y}^{(i)}-y^{(i)})^2=\ \frac{1}{8}\sum_{i=1}^{4}(\hat{y}^{(i)}-y^{(i)})^2=$$

$$\frac{1}{8}\sum_{i=1}^{4}(\hat{y}^{(i)}-y^{(i)})^2=\ \frac{1}{8}\ [\,(17.15-37.99)^2+(26.00-47.34)^2+(25.55-44.38)^2+(13.40-28.17)^2\,]=\ 182.8028$$

$$MAE=\frac{1}{m}\sum_{i=1}^{m}\left|\hat{y}^{(i)}-y^{(i)}\right|=\ \frac{1}{8}\sum_{i=1}^{4}\left|\hat{y}^{(i)}-y^{(i)}\right|=$$

$$\frac{1}{4}\sum_{i=1}^{4}\left|\hat{y}^{(i)}-y^{(i)}\right|=\ \frac{1}{4}\ [\,|17.15-37.99|+|26.00-47.34|+|25.55-44.38|+|13.40-28.17|]=\ 18.945$$

ب)

$$L(w) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 = \frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$L_i(w) = \frac{1}{2}(\hat{y}^{(i)} - y^{(i)})^2$$

$$\frac{\partial L_i(w)}{\partial w_j} = \left(\hat{y}^{(i)} - y^{(i)}\right)x_j^{(i)}\ for\ i = 1.2$$

$$\frac{\partial L_i(w)}{\partial w_0} = \left(\hat{y}^{(i)} - y^{(i)}\right)$$

`w0, w1, w2 =` [-59.5, -0.15, 0.6]

داده سوم انتخاب شده:

$$\frac{\partial L_i(w)}{\partial w_1} = (25.55 - 44.38) * 37 = -696.71$$

$$\frac{\partial L_i(w)}{\partial w_2} = (25.55 - 44.38) * 151 = -2843.33$$

$$\frac{\partial L_i(w)}{\partial w_0} = (25.55 - 44.38) = -18.83$$

$w[0] = -59.50 - 0.1 * (-18.83\ ) = -57.617$

$w[1] = -0.15 - 0.1 * (-696.71\ ) = 69.521$

$w[2] = 0.60 - 0.1 * (-2843.33) = 284.933$

$\hat{y}^{(1)} = w[0] + w[1] * 41 + w[2] * 138 = -57.617 + 69.521 * 41 + 284.933 * 138$
$= 42113.498$

داده اول انتخاب شده:

$$\frac{\partial L_i(w)}{\partial w_1} = (42113.498 - 37.99) * 41 = 1725095.828$$

$$\frac{\partial L_i(w)}{\partial w_2} = (42113.498 - 37.99) * 138 = 5806420.104$$

$$\frac{\partial L_i(w)}{\partial w_0} = (42113.498 - 37.99) = 42075.508$$

$w[0] = -57.617 - 0.1 * (42075.508\ ) = -4265.1678$

$w[1] = 69.521 - 0.1 * (1725095.828\ ) = -172440.0618$

$w[2] = 284.933 - 0.1 * (5806420.104) = -580357.0774$

$$\hat{y}^{(1)} = w[0] + w[1] * 41 + w[2] * 138 = -4265.1678 + -172440.0618 * 41 + -580357.0774 * 138 = -87163584.3828$$

$$\hat{y}^{(2)} = w[0] + w[1] * 42 + w[2] * 153 = -4265.1678 + -172440.0618 * 42 + -580357.0774 * 153 = -96041380.6056$$

$$\hat{y}^{(3)} = w[0] + w[1] * 37 + w[2] * 151 = -4265.1678 + -172440.0618 * 37 + -580357.0774 * 151 = -94018466.1418$$

$$\hat{y}^{(4)} = w[0] + w[1] * 46 + w[2] * 133 = -4265.1678 + -172440.0618 * 46 + -580357.0774 * 133 = -85123999.30479999$$

$$MSE = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 =$$

$$\frac{1}{8}\sum_{i=1}^{4}(\hat{y}^{(i)} - y^{(i)})^2 = \frac{1}{8}\,[(-87163584.3828 - 37.99)^2 + (-96041380.6056 - 47.34)^2 + (-94018466.1418 - 44.38)^2 + (-85123999.30479999 - 28.17)^2] = 4113379165155784.5$$

ج)

$$J(w_0.w_1.w_2) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 = \frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial J(w_0.w_1.w_2)}{\partial w_j} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)x_j^{(i)} for\ j = 1.2$$

$$\frac{\partial J(w_0.w_1.w_2)}{\partial w_0} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)$$

$$\frac{\partial J(w_0.w_1.w_2)}{\partial w_0} = \frac{1}{4}\sum_{i=1}^{4}(\hat{y}^{(i)} - y^{(i)}) = \frac{1}{4}\,[(17.15 - 37.99) + (26.00 - 47.34) + (25.55 - 44.38) + (13.40 - 28.17)] = -18.945$$

$$\frac{\partial J(w_0.w_1.w_2)}{\partial w_1} = \frac{1}{4}\sum_{i=1}^{4}(\hat{y}^{(i)} - y^{(i)})x_1^{(i)} = \frac{1}{4}\,[(17.15 - 37.99) * 41 + (26.00 - 47.34) * 42 + (25.55 - 44.38) * 37 + (13.40 - 28.17) * 46\,] = -781.7125$$

$$\frac{\partial J(w_0.w_1.w_2)}{\partial w_2} = \frac{1}{4}\sum_{i=1}^{4}(\hat{y}^{(i)} - y^{(i)})x_2^{(i)} = \frac{1}{4}\,[(17.15 - 37.99) * 138 + (26.00 - 47.34) * 153 + (25.55 - 44.38) * 151 + (13.40 - 28.17) * 133\,] = -2737.17$$

$$w[0] = -59.50 - 0.1 * (-18.945\ ) = \ -57.6055$$

$$w[1] = \ -0.15 - 0.1 * (-781.7125\ ) = 78.02125$$

$$w[2] = 0.60 - 0.1 * (-2737.17\ ) = 274.317$$

$$\hat{y}^{(1)} = \ w[0] + w[1] * 41 + w[2] * 138 = \ -57.6055 + \ 78.02125 * 41 + 274.317 * 138 = 40997.01175$$

$$\hat{y}^{(2)} = \ w[0] + w[1] * 42 + w[2] * 153 = \ -57.6055 + \ 78.02125 * 42 + 274.317 * 153 = \ 45189.788$$

$$\hat{y}^{(3)} = \ w[0] + w[1] * 37 + w[2] * 151 = \ -57.6055 + \ 78.02125 * 37 + 274.317 * 151 = 44251.04775$$

$$\hat{y}^{(4)} = \ w[0] + w[1] * 46 + w[2] * 133 = \ -57.6055 + \ 78.02125 * 46 + 274.317 * 133 = 40015.5334$$

$$MSE = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 = \frac{1}{8}\sum_{i=1}^{4}(\hat{y}^{(i)} - y^{(i)})^2 =$$

$$\frac{1}{8}\sum_{i=1}^{4}(\hat{y}^{(i)} - y^{(i)})^2 = \ \frac{1}{8}\ [\,(40997.01175 - 37.99)^2 + (45189.788 - 47.34)^2 + (44251.04775 - 44.38)^2 + (40015.533 - 28.17)^2 = 908587593.4252921$$

سوال ۴)

الف) در نقاط $x = -\sqrt{2}$ و $x = \sqrt{2}$ به سبب عدم وجود مشتق از مشتق راست و چپ به عنوان مشتق تابع استفاده می کنیم و خواهیم داشت:

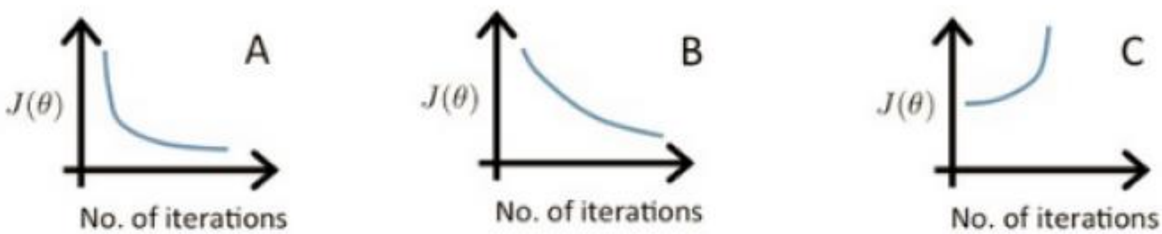
$$f(x) = \frac{1}{2}x^2 . f'(x) = x\ for - \sqrt{2} < x < \sqrt{2}$$

$$f'(\sqrt{2}) = \sqrt{2}\ and\ f'(-\sqrt{2}) = -\sqrt{2}$$

ب) به ازای تمام نقاطی که از $\sqrt{2}$ بزرگتر یا مساوی هستند یا از $\sqrt{2}$ - کوچکتر یا مساوی هستند از ***subgradient*** فوق بهره گیری می نماییم. و لذا برای x های بزرگتر و یا مساوی $\sqrt{2}$ مقدار ***subgradient*** برابر $\sqrt{2}$ و برای نقاط کوچکتر یا مساوی $\sqrt{2}$ - مقدار ***subgradient*** برابر $\sqrt{2}$ - است. پس در همه نقاط یکسان نمی باشد اما برای نقاط موجود در هر یک از دو بازه ذکر شده یکسان است.

ج) در محاسبه تابع هزینه با بهره گیری از ***MAE*** نیاز به استفاده از ***subgradient*** داریم چراکه در نقاطی که مقدار پیش بینی شده با مقدار واقعی داده یکسان باشد به نقاط مشتق ناپذیر برخورد کرده و می بایست در این نقاط از ***subgradient*** استفاده کنیم.

سوال ۵)

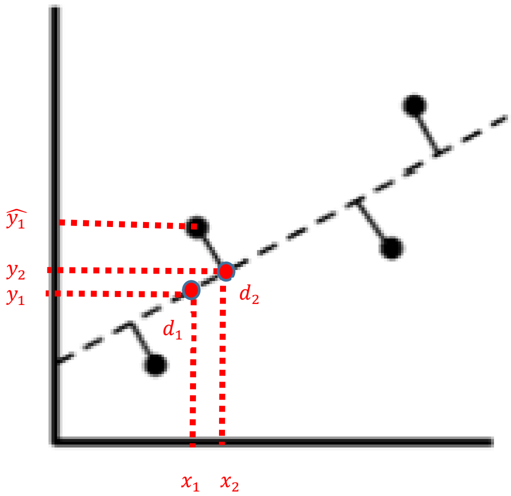


می دانیم در صورتی که نرخ یادگیری بیش از حد بزرگ باشد تابع هزینه نه تنها همگرا نخواهد شد بلکه روند صعودی پیدا کرده و در طی گام های متوالی افزایش می یابد با توجه به این امر می توان گفت چون نمودار B و A هر دو روند نزولی دارند اما نمودار C یک روند صعودی در پیش گرفته است لذا نمودار C متعلق به یادگیری داده ها با نرخ یادگیری نسبتا بزرگ و بزرگ تر از نرخ یادگیری دو نمودار A و B است. از میان دو نمودار A و B شیب نمودار A بیشتر بوده و سرعت همگرایی در این نمودار نیز بیشتر است در حالی که در نمودار B سرعت همگرایی کمتر بوده و شیب نمودار نیز مایل تر و به فرم نمودار خطی نزدیک تر است از طرفی می دانیم هر چه شیب نمودار کمتر و به نمودار خطی نزدیک تر باشد حاکی از نرخ یادگیری کوچکتری می باشد لذا از میان دو نمودار B و A می توان گفت نرخ یادگیری در نمودار B کوچکتر از نرخ یادگیری در نمودار A بوده است.

نتیجه کلی : Learning Rate B < Learning Rate A < Learning Rate C

سوال ۶)

با توجه به اینکه هدف از پیاده سازی تابع هزینه بررسی میزان خطای پیش بینی نسبت به مقدار واقعی داده ها است می بایست از vertical offset استفاده نماییم چرا که برای مثال اگر فرض کنیم نمودار زیر نمودار قیمت خانه برحسب یک پارامتر مانند متراژ خانه باشد آنگاه مقدار قیمت پیش بینی شده برای خانه ای مانند d_1 که متراژ آن x_1 است برابر y_1 می باشد این در حالی است که قیمت واقعی این خانه برابر $\widehat{y_1}$ می باشد و لذا تابع هزینه می بایست تابعی از اختلاف میان y_1 و $\widehat{y_1}$ باشد. در حالی که y_2 قیمت پیش بینی شده برای خانه دیگری مانند d_2 است که متراژ آن x_2 می باشد و چنانچه تابع هزینه تابعی از اختلاف میان y_2 و $\widehat{y_1}$ باشد در حقیقت تابعی از اختلاف قیمتهای پیش بینی شده برای خانه d_1 و d_2 است و نه تابعی از اختلاف قیمت واقعی و قیمت پیش بینی شده برای یک خانه همانند d_1.



سوال ۷)

الف) در شرایطی که داده های یادگیری شامل داده های پرت باشند استفاده از MAE به جای MSE مناسب تر است چراکه توان ۲ در تابع هزینه MSE منجر به افزایش چشمگیر مقدار تابع می شود در حالتی که این تاثیر در MAE تنها به اندازه اختلاف دو مقدار پیش بینی شده و مقدار داده پرت است و توان دوم از آن نیست.

ب) به این سبب که در بخش زیادی از داده ها $y_n = \widehat{y_n}$ است لذا استفاده از تابع هزینه MAE منجر به افزایش تعداد نقاط مشتق ناپذیر گشته و به دلیل لزوم استفاده از subgradient بهتر است از MSE به عنوان تابع هزینه استفاده نماییم تا با نقاط مشتق ناپذیر مواجه نشویم.