



دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف اول درس مبانی یادگیری ماشین (بخش تئوری)

نام و نام خانوادگی: علیرضا ابره فروش

شماره دانشجویی: ۹۸۱۶۶۰۳

نیم سال تحصیلی: پاییز ۱۴۰۱

مدرس: دکتر مهران صفایانی

۱

۱.۱ الف

- مسیر آبی که نسبتاً مستقیم است مربوط به روش Batch GD است. چون در هر گام کل مجموعه‌ی داده‌ها را می‌بیند.
- مسیر قرمز که نوسان بیشتری نسبت به مسیر آبی و نوسان کمتری نسبت به مسیر سبز دارد مربوط به روش Mini-batch GD است. چون در هر گام بخشی از مجموعه‌ی داده‌ها را می‌بیند.
- مسیر سبز که در مقایسه با بقیه مسیرها بیشترین نوسان را دارد مربوط به روش SGD است. چون در هر گام یک داده از مجموعه داده‌ها را می‌بیند.

۲.۱ ب

داده‌های تصویر a برخلاف داده‌های تصویر b نرمال نشده‌اند. در داده‌هایی که featureها scale نشده‌اند، تغییر مقادیر پارامترها باعث بروز اعوجاج شدیدتر نسبت به داده‌های نرمال شده می‌شود و در نتیجه همگرایی مدل را کند می‌کند.

۲

۱.۲ الف

MAE بهتر است. چون تعداد داده‌های outlier قابل توجه است، در صورت استفاده از MSE، توان دوی featureهای این داده‌ها اثر داده می‌شوند و اثر سایر داده‌ها (که outlier نیستند و مدل را مطلوب‌تر آموزش می‌دهند) تقریباً ناچیز می‌شود.

۲.۲ ب

MSE بهتر است. چون احتمالاً مدل دچار overfitting شده است، مقادیر پیش‌بینی شده به مقادیر واقعی خیلی نزدیک‌اند و MSE آن‌ها را بیشتر بازتاب می‌دهد.

۳.۲ ج

MAE بهتر است. از آنجایی که داده‌ها مقیاس‌های متفاوتی دارند، اگر از MSE استفاده کنیم اثر نرمال نبودن داده‌ها به طور نمایی در مدل تاثیر می‌گذارد.

۳

ستون A که به نسبت سایر ستون‌ها ضرایب بزرگتری دارد فاقد regularization term است. پس مربوط به تابع هزینه‌ی اول است.
 ستون B فاقد ضرایب صفر است. می‌دانیم که امکان صفر شدن ضرایب در lasso regression موجود است. پس مربوط به تابع دوم است.
 ستون C دارای ضرایب صفر است. می‌دانیم که در ridge regression هرگز ضرایب صفر نمی‌شوند. پس مربوط به تابع سوم است.

۴

$$p(x_k|\theta) = \sqrt{\theta} x_k^{\sqrt{\theta}-1}$$

$$p(X|\theta) = \prod_{k=1}^n p(x_k|\theta)$$

$$\ln(p(X|\theta)) = \sum_{k=1}^n \ln(p(x_k|\theta)) = \sum_{k=1}^n \ln(\sqrt{\theta} x_k^{\sqrt{\theta}-1}) = \sum_{k=1}^n \left[\frac{1}{2} \ln(\theta) + (\sqrt{\theta} - 1) \ln(x_k) \right]$$

$$\frac{\partial}{\partial \theta} \ln(p(X|\theta)) = \frac{n}{2\theta} - \frac{\sum_{k=1}^n \ln(x_k)}{2\sqrt{\theta}} = 0$$

$$\theta = \left(\frac{n}{\sum_{k=1}^n \ln(x_k)} \right)^2$$

۵

$$p(\mu|X) \propto p(x_k|\mu) \cdot p(\mu)$$

$$p(\mu|X) = \left[\prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\frac{(x_k-\mu)^2}{2\sigma'^2}} \right] \cdot \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_\mu^2} e^{-\frac{\|\mu-\mu_0\|^2}{2\sigma_\mu^2}}$$

$$\ln(p(\mu|X)) = \sum_{k=1}^n \left[-\ln(\sqrt{2\pi\sigma'^2}) - \frac{(x_k-\mu)^2}{2\sigma'^2} \right] - \ln((2\pi)^{\frac{1}{2}} \sigma_\mu^2) - \frac{\|\mu-\mu_0\|^2}{2\sigma_\mu^2}$$

$$\frac{\partial}{\partial \mu} \ln(p(\mu|X)) = 0$$

$$\sum_{k=1}^n \frac{x_k - \mu}{\sigma'^2} = \frac{\|\mu - \mu_0\|}{2\sigma_\mu^2}$$

$$\mu = \frac{\frac{\sum_{k=1}^n x_k}{\sigma'^2} + \frac{\mu}{2\sigma_\mu^2}}{n + \frac{1}{2\sigma_\mu^2}}$$

۶

 $\mathcal{L}(w)$: تابع هزینه $\mathcal{L}_i(w)$: هزینهی training example i ام w^t : وزن‌ها در گام t ام

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_9 \end{bmatrix}, X = \begin{bmatrix} (X_1)^2 & X_1 & 1 \\ (X_2)^2 & X_2 & 1 \\ \vdots & \vdots & \vdots \\ (X_9)^2 & X_9 & 1 \end{bmatrix}, w = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\mathcal{L}_i(w^{(t)}) = \frac{1}{2} (y_i - X_i^T w^{(t)})^2$$

$$\nabla \mathcal{L}_i(w^{(t)}) = -X_i (y_i - X_i^T w^{(t)})$$

$$w^{(t+1)} = w^{(t)} - \alpha \nabla \mathcal{L}_i(w^{(t)})$$

$$w^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$w^{(1)} = \begin{bmatrix} 0.00000000e+00 \\ 0.00000000e+00 \\ 0.00000000e+00 \end{bmatrix} - 0.1 \times \left(- \begin{bmatrix} 35.38^2 \\ 35.38 \\ 1 \end{bmatrix} \left(2955.53 - \begin{bmatrix} 35.38^2 & 35.38 & 1 \end{bmatrix} \begin{bmatrix} 0.00000000e+00 \\ 0.00000000e+00 \\ 0.00000000e+00 \end{bmatrix} \right) \right) =$$

$$\begin{bmatrix} 3.69956813e+05 \\ 1.04566651e+04 \\ 2.95553000e+02 \end{bmatrix}$$

$$w^{(2)} = \begin{bmatrix} 3.69956813e+05 \\ 1.04566651e+04 \\ 2.95553000e+02 \end{bmatrix} - 0.1 \times \left(- \begin{bmatrix} 15.32^2 \\ 15.32 \\ 1 \end{bmatrix} \left(560.30 - \begin{bmatrix} 15.32^2 & 15.32 & 1 \end{bmatrix} \begin{bmatrix} 3.69956813e+05 \\ 1.04566651e+04 \\ 2.95553000e+02 \end{bmatrix} \right) \right) =$$

$$\begin{bmatrix} -2.04129879e+09 \\ -1.33257738e+08 \\ -8.69867277e+06 \end{bmatrix}$$

$$w^{(3)} = \begin{bmatrix} -2.04129879e+09 \\ -1.33257738e+08 \\ -8.69867277e+06 \end{bmatrix} - 0.1 \times \left(- \begin{bmatrix} 11.74^2 \\ 11.74 \\ 1 \end{bmatrix} \left(334.32 - \begin{bmatrix} 11.74^2 & 11.74 & 1 \end{bmatrix} \begin{bmatrix} -2.04129879e+09 \\ -1.33257738e+08 \\ -8.69867277e+06 \end{bmatrix} \right) \right) =$$

$$\begin{bmatrix} 3.89738346e+12 \\ 3.32015359e+11 \\ 2.82833471e+10 \end{bmatrix}$$

$$w^{(4)} = \begin{bmatrix} 3.89738346e+12 \\ 3.32015359e+11 \\ 2.82833471e+10 \end{bmatrix} - 0.1 \times \left(- \begin{bmatrix} 19.05^2 \\ 19.05 \\ 1 \end{bmatrix} \left(864.44 - \begin{bmatrix} 19.05^2 & 19.05 & 1 \end{bmatrix} \begin{bmatrix} 3.89738346e+12 \\ 3.32015359e+11 \\ 2.82833471e+10 \end{bmatrix} \right) \right) =$$

$$\begin{bmatrix} -5.15545092e+16 \\ -2.70614602e+15 \\ -1.42044054e+14 \end{bmatrix}$$

$$\begin{aligned}
w^{(5)} &= \begin{bmatrix} -5.15545092e + 16 \\ -2.70614602e + 15 \\ -1.42044054e + 14 \end{bmatrix} - 0.1 \times \begin{bmatrix} 26.85^2 \\ 26.85 \\ 1 \end{bmatrix} \left(1709.09 - \begin{bmatrix} 26.85^2 & 26.85 & 1 \end{bmatrix} \begin{bmatrix} -5.15545092e + 16 \\ -2.70614602e + 15 \\ -1.42044054e + 14 \end{bmatrix} \right) = \\
&\begin{bmatrix} 2.68463555e + 21 \\ 9.99856406e + 19 \\ 3.72381873e + 18 \end{bmatrix} \\
w^{(6)} &= \begin{bmatrix} 2.68463555e + 21 \\ 9.99856406e + 19 \\ 3.72381873e + 18 \end{bmatrix} - 0.1 \times \begin{bmatrix} 39.45^2 \\ 39.45 \\ 1 \end{bmatrix} \left(3670.48 - \begin{bmatrix} 39.45^2 & 39.45 & 1 \end{bmatrix} \begin{bmatrix} 2.68463555e + 21 \\ 9.99856406e + 19 \\ 3.72381873e + 18 \end{bmatrix} \right) = \\
&\begin{bmatrix} -6.50851298e + 26 \\ -1.64980998e + 25 \\ -4.18201594e + 23 \end{bmatrix} \\
w^{(7)} &= \begin{bmatrix} -6.50851298e + 26 \\ -1.64980998e + 25 \\ -4.18201594e + 23 \end{bmatrix} - 0.1 \times \begin{bmatrix} 30.51^2 \\ 30.51 \\ 1 \end{bmatrix} \left(2202.93 - \begin{bmatrix} 30.51^2 & 30.51 & 1 \end{bmatrix} \begin{bmatrix} -6.50851298e + 26 \\ -1.64980998e + 25 \\ -4.18201594e + 23 \end{bmatrix} \right) = \\
&\begin{bmatrix} 5.64425427e + 31 \\ 1.84997346e + 30 \\ 6.06351097e + 28 \end{bmatrix} \\
w^{(8)} &= \begin{bmatrix} 5.64425427e + 31 \\ 1.84997346e + 30 \\ 6.06351097e + 28 \end{bmatrix} - 0.1 \times \begin{bmatrix} 3.98^2 \\ 3.98 \\ 1 \end{bmatrix} \left(13.08 - \begin{bmatrix} 3.98^2 & 3.98 & 1 \end{bmatrix} \begin{bmatrix} 5.64425427e + 31 \\ 1.84997346e + 30 \\ 6.06351097e + 28 \end{bmatrix} \right) = \\
&\begin{bmatrix} -1.37156316e + 33 \\ -3.56945428e + 32 \\ -9.00889632e + 31 \end{bmatrix} \\
w^{(9)} &= \begin{bmatrix} -1.37156316e + 33 \\ -3.56945428e + 32 \\ -9.00889632e + 31 \end{bmatrix} - 0.1 \times \begin{bmatrix} 0.29^2 \\ 0.29 \\ 1 \end{bmatrix} \left(2.28 - \begin{bmatrix} 0.29^2 & 0.29 & 1 \end{bmatrix} \begin{bmatrix} -1.37156316e + 33 \\ -3.56945428e + 32 \\ -9.00889632e + 31 \end{bmatrix} \right) = \\
&\begin{bmatrix} -1.36896487e + 33 \\ -3.47985832e + 32 \\ -5.91938034e + 31 \end{bmatrix}
\end{aligned}$$

در هر گام از بین training example ها یکی را به صورت تصادفی انتخاب می کنیم. این کار را به تعداد training example ها تکرار می کنیم تا کل داده ها توسط مدل دیده شوند.

منابع