علی ترابی پور - ۹۶۱۰۵۹۹۴

تمرین چهارم ML

دکتر حسینی.

Closed form solution for regression:

* least_squares : $\|y - X.w\|^2$, $w^* = (X^T.X)^{-1}.X^T.y$

$\|y - Xw\|^2 = LSE \longrightarrow \nabla_w LSE = -2.X^T(y - Xw) = 0$ مشتق می‌گیریم

$\longrightarrow X^T.y = X^T.X.w \longrightarrow \boxed{w^* = (X^T X)^{-1}.X^T.y}$ ← w بهینه

* ridge regression: $\|y - Xw\|^2 + \lambda.\|w\|^2$, $w^* = ?$

$\nabla_w \text{ridge} = -2X^T(y - Xw) + 2\lambda.w = 0 \longrightarrow -X^T y + X^T X w$

$+ \lambda w = 0 \longrightarrow (X^T X + \lambda I)w = X^T.y \longrightarrow \boxed{w^* = (X^T X + \lambda I)^{-1}.X^T y}$

Regression and Gradient descent:

$$* \quad y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + \varepsilon \qquad \text{where } \varepsilon \sim N(0, \sigma^2)$$

$$P(y \mid x_1, x_2) = P(w^T X + \varepsilon \mid X) \simeq N(w^T X, \sigma^2)$$

$$* \quad P(y \mid X, w) = \prod_{i=1}^{n} P(y_i \mid x_1^i, x_2^i, w) = \prod_{i=1}^{n} N(y \mid w^T X, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \times e^{-\frac{(y^i - w^T x^i)^2}{2\sigma^2}} \xrightarrow[\text{Constant } X]{\ln()} -n\sigma - \frac{n}{2}\ln(2 w^T X) -$$

$$\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (y^i - w^T x^i)^2 \longrightarrow \max P(y \mid X, w) \sim \min \sum_{i=1}^{n} (y^i - w^T x^i)^2$$

$* \quad$ تابع هزینه بخش قبل : $f(w) = \sum_{i=1}^{n} (y^i - w^T x^i)^2$

$$* \quad \frac{\partial f}{\partial w_0} = -2 \sum_{i=1}^{n} (y^i - w^T x^i) \qquad \frac{\partial f}{\partial w_1} = -2 \sum_{i=1}^{n} (y^i - w^T x^i)(x_1^i)$$

$$\frac{\partial f}{\partial w_2} = -2 \sum_{i=1}^{n} (y^i - w^T x^i)(x_2^i) \qquad \frac{\partial f}{\partial w_3} = -2 \sum_{i=1}^{n} (y^i - w^T x^i)(x_1^i)^2$$

$$\frac{\partial f}{\partial w_4} = -2 \sum_{i=1}^{n} (y^i - w^T x^i)(x_2^i)^2$$

$$\Rightarrow \nabla_w f = \left[\frac{\partial f}{\partial w_0}, \frac{\partial f}{\partial w_1}, \ldots, \frac{\partial f}{\partial w_4} \right] \Leftarrow$$

گرادیان کاهشی $\longrightarrow$ $\boxed{w_{n+1} = w_n - \alpha \cdot \nabla_w f(w)}$

ج	پ	چ	س	د	ی	ش
۳	۲	۱				
۱۰	۹	۸	۷	۶	۵	۴
۱۷	۱۶	۱۵	۱۴	۱۳	۱۲	۱۱
۲۴	۲۳	۲۲	۲۱	۲۰	۱۹	۱۸
	۳۰	۲۹	۲۸	۲۷	۲۶	۲۵

Prababilistic modeling :

* $P(y|x;w,\beta) = f(x,w) + \varepsilon$ where $\varepsilon \sim N(0,\beta^{-1})$

maximize log-liklihood $\equiv$ minimize Sum square error

$t = y(x,w) + \varepsilon \rightsquigarrow P(t|y(x,w)) \equiv N(y(x,w),\beta^{-1})$

$t = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad P(t_{1:N}|x_{1:N},w) = \prod_{n=1}^{N} P(t_i|x_i,w)$

$= \prod_{i=1}^{N} N(t_i|\underbrace{w^T\phi(x)}_{y(x,w)},\beta^{-1}) \rightsquigarrow w^* = \arg\max_w P(t_{1:N},w)$

$= \arg\max_w \ln P(t_{1:N}|x_{1:N},w) = \sum_{i=1}^{N} \ln \underbrace{P(t_i|x_i,w)}_{N(w^T\phi(x),\beta^{-1})} =$

$\frac{\beta^{\frac{1}{2}}}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(t-w^T\phi(x))^2}{2}\right)$

$= \frac{N}{2}\ln\beta - \beta\sum_{i=1}^{N}\frac{(t-y(x,w))^2}{2} \rightsquigarrow w^* = \arg\min_w \frac{\beta}{2}\sum_{i=1}^{N}(t_n - y(x_n,w))^2$

$\longrightarrow$ MLLE $\equiv$ MSE ✓

* Prior: $P(w;\alpha) = N(0, \alpha^{-1} I) \xrightarrow[MAP]{Bayes}$ ridge_reg Sum Square

$$P(w) = \prod_{i=0}^{M} P(w_i) = \prod_{i=0}^{M} N(w_i|0, \alpha^{-1})$$

$$P(D|w) = \prod_{i=1}^{N} p(t_i|x_i, w)$$

$$P(w|D) = \frac{P(D|w)P(w)}{P(D)} \rightsquigarrow w^*_{MAP} = \arg\min -\ln P(D|w) - \ln P(w)$$

$$= \arg\min -\sum_{n=1}^{N} \ln N(t_n|x_n, w_n) - \sum_{i=0}^{M} \ln P(w_i) \qquad \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\alpha}{2} w_i^2}$$

$$= \arg\min_w \frac{\beta}{2} \sum_{n=0}^{N} (t_n - y_n(x_n, w))^2 + \frac{\alpha}{2} \sum_{i=0}^{M} w_i^2$$

$$= \arg\min_w \underbrace{\sum_{n=1}^{N} (t_n - y(x_n, w))^2}_{\text{LSE-term}} + \underbrace{\underbrace{\frac{\alpha}{\beta}}_{\lambda} \sum_{i=0}^{M} w_i^2}_{\text{regularication-term}}$$

✓

$$*\, P(w|D) = \frac{P(D|w)\,P(w)}{P(D)} \qquad P(w) = L(0, \alpha^{-1} I)$$

$$\rightarrow P(w) = \prod_{i=1}^{n} \frac{\sqrt{2\alpha}}{2} \exp\left(\frac{-|w_i - 0|}{(2\alpha)^{-\frac{1}{2}}} \right)$$

$$w_{map} = \arg\max_w \left(\log P(D|w) + \log P(w) \right) = \arg\max_w \Big(\frac{N}{2} \log\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2} \sum_{i=1}^{n} \left(y(x^i, w) - t^i \right)^2 + N \log \frac{\sqrt{2\alpha}}{2} - \sqrt{2\alpha} \sum_{i=1}^{n} |w_i| \Big)$$

$$= \arg\min_w \Big(\underbrace{\sum_{i=1}^{n} \left(t^i - y(x^i, w) \right)^2}_{LSE} + \underbrace{\sum_{i=1}^{n} |w_i|}_{\text{lasso-regularization term}} \Big)$$

با توجه به دو بخش اخیر می‌توان متوجه شد Prior های مختلف منجر به term های متفاوت

regularization برای تابع خطای مورد نظر تعبیه شدن می‌شوند.