Assignment 4: Stochastic Linear Regression

Haoyang Han

Problem 1: prove posterior distribution over the parameters $\tilde{w}$:

$$Pr(\tilde{w}|\tilde{X},\tilde{y}) = Norm_w[\sigma^{-2}A^{-1}\tilde{X}\tilde{y}, \tilde{A}^{-1}]$$

$$A = \sigma^{-2}\tilde{X}\tilde{X}^T + \sigma_p^{-2}\vec{I}$$

Prof: Since Bayes theorem is:

$$P(X,y) = P(Y/X)\cdot P(X)$$

$$\therefore P(y/X) = \frac{P(X/y)\cdot P(y)}{P(X)} \quad (*)$$

and for a given dataset $\{(\tilde{x}_p, y_p)\}_{p=1}^{P}$, have following condition:

It's likelihood: $Pr(\bar{y}/\tilde{x},\tilde{w}) = Norm_y[\tilde{x}^T\tilde{w}, \sigma^2 I]$

and its distribution on $\tilde{w}$: $Pr(\tilde{w}) = Norm_{\tilde{w}}[0, \sigma_p^2 I]$

Our goal: to get Posterior: $Pr(\tilde{w}/\tilde{X},\bar{y})$

using Bayes theorem: $Pr(\bar{y}/\tilde{x},\tilde{w})\cdot Pr(\tilde{w}) / Pr(\bar{y}/\tilde{x})$

Besides,

$$\begin{cases} p(x) = Norm_x(\mu, \Lambda^{-1}) \\ p(y|x) = Norm_y(Ax+b, L^{-1}) \\ p(y) = Norm_y(Ax+b, L^{-1}+A\Lambda^{-1}A^T) \end{cases}$$

$$\therefore P(x|y) = Norm_x(\Sigma(A^T L(y-b)+\Lambda\mu), \Sigma)$$

in which $\Sigma = (\Lambda + A^T L A)^{-1}$, $A^T = \tilde{x}$, $L = \sigma^{-2} I$, $\Lambda = \sigma_p^{-2} I$, $\mu = 0$, $b = 0$

$$\therefore Pr(\tilde{w}/\tilde{x},\bar{y}) = Pr(\bar{y}/\tilde{x},\tilde{w})\cdot Pr(\tilde{w}) / Pr(\bar{y}/\tilde{x})$$

$$= Norm_{\tilde{w}}[\tfrac{1}{\sigma^2}B^{-1}\tilde{x}\bar{y}, B^{-1}], \quad B = \tfrac{1}{\sigma^2}\tilde{x}\tilde{x}^T + \tfrac{1}{\sigma_p^2} I$$

Rewrite formula:

$$Pr(\tilde{w}/\tilde{x},\bar{y}) = Norm_{\tilde{w}}[\sigma^{-2}A^{-1}\tilde{X}\bar{y}, A^{-1}], \text{ where } A = \sigma^{-2}\tilde{x}\tilde{x}^T + \sigma_p^{-2} I$$

Q.E.D.

Problem 2: Draw inference

Prof:

for distribution of new dataset $\{x_i^*, y_i^*\}$ based on $\{x, y\}$

$\bar{x}^* \rightarrow Pr(y^* / \bar{x}^*, \tilde{x}, \bar{y})$

Prove $Pr(y^* | \bar{x}^*, \tilde{X}, \bar{y}) = Norm_{y^*}[\sigma^{-2} \bar{x}^{*T} A^{-1} \tilde{X} \bar{y}, \bar{x}^{*T} A^{-1} x^* + \sigma^2]$

Sol: $\because Pr(y^* / \bar{x}^*, \tilde{X}, \bar{y})$

$= \int Pr(y^* / \bar{x}^*, \tilde{w}) \, Pr(\tilde{w} / \tilde{x}, \bar{y}) \, dw$ (for all possible w)

from Problem 1: $\begin{cases} Pr(\bar{y}^* / \tilde{x}^*, \tilde{w}) = Norm_{\bar{y}}[\tilde{x}^* \tilde{w}, \sigma^2 I] \\ Pr(\tilde{w} / \tilde{x}, \bar{y}) = Norm_{\tilde{w}}[\sigma^{-2} A^{-1} \tilde{X} \bar{y}, A^{-1}] \end{cases}$

$\Rightarrow Pr(y^* / \bar{x}^*, \tilde{X}, \bar{y}) = \int Norm_{y^*}[\tilde{x}^{*T} \tilde{w}, \sigma^2 I] \cdot Norm_{\tilde{w}}[\sigma^{-2} A^{-1} \tilde{X} \bar{y}, A^{-1}] dw$

$= Norm_{y^*}[\sigma^{-2} \bar{x}^{*T} A^{-1} \tilde{X} \bar{y}, \bar{x}^{*T} A^{-1} \bar{x}^* + \sigma^2]$

in which $\begin{cases} \tilde{x}^* = \tilde{x}_{p+1} \\ \bar{y}^* = \bar{y}_{p+1} \end{cases}$

Another approach: calculate mean / var

$\because \begin{cases} y^* = (x^*)^T \cdot \tilde{w} + \varepsilon_i \ (w.N.) \\ \tilde{w} = \sigma^{-2} A^{-1} X w + \eta_i \ (w.N.) \end{cases}$

$\therefore E(y^*) = (x^*)^T \cdot E(w)$

$= \sigma^{-2} x^{*T} A^{-1} X w$

$E(y^{*2}) = (x^*)^T E(w w^T) x^* + E(\varepsilon_i \varepsilon_i^T)$

$= (x^*)^T E(w w^T) x^* + \sigma^2$

$= (x^*)^T [A^{-1} + \frac{1}{\sigma^4} A^{-1} X w w^T X^T A^{-1}] x^* + \sigma^2$

$= (x^*)^T A^{-1} x^T + \sigma^2$

$\therefore Pr(y^* / \bar{x}^*, \tilde{X}, \bar{y}) = Norm_{y^*}[\sigma^{-2} \bar{x}^{*T} A^{-1} \tilde{X} \bar{y}, \bar{x}^{*T} A^{-1} \bar{x}^* + \sigma^2]$

Problem 3: Logistic Regression:

Prof: Given $\{(\vec{x}_i, y_i)\}$ with $y_i \in$ Bernoulli distribution

$$\begin{cases} Pr(x=0) = (1-\lambda) \\ Pr(x=1) = \lambda \end{cases}$$

$$\therefore Pr(x) = \lambda^{x}(1-\lambda)^{1-x} = Bern_x[\lambda],$$

Let $\lambda = b + \vec{x}^T \vec{w}$

and $Pr(y_i / b, \vec{w}, \vec{x}_i) = Bern_{y_i}[sig(b + \vec{w}^T \vec{x}_i)]$

assume datasets independence, Log posterior probability.

$$L = \sum_{i=1}^{p} y_i \log(sig(a_i)) + \sum_{i=1}^{p} (1-y_i) \log(1 - sig(a_i))$$

$$= \sum_{i=1}^{p} y_i \log \frac{1}{1+e^{-\vec{x}_i^T \vec{w}}} + \sum_{i=1}^{p} (1-y_i) \log\left(\frac{e^{-\vec{x}_i^T \vec{w}}}{1+e^{-\vec{x}_i^T \vec{w}}}\right)$$

$$\therefore \frac{\partial}{\partial w_j} L = \sum_{i=1}^{p} y_i \frac{1}{sig(a_i)} \frac{\partial}{\partial w_j}\left(\frac{1}{1+e^{-\vec{x}_i^T \vec{w}}}\right) + \sum_{i=1}^{p} (1-y_i) \frac{1}{1-sig(a_i)}\left[-\frac{\partial}{\partial w_j}\left(\frac{1}{1+e^{-\vec{x}_i^T \vec{w}}}\right)\right]$$

$$= \sum_{i=1}^{p} y_i (-sig(a_i)) (e^{-\vec{x}_i^T \vec{w}})(-\vec{x}_{ij}^T)$$

$$+ \sum_{i=1}^{p} (1-y_i) \frac{1+e^{a_i}}{e^{-a_i}} \left(-\frac{1}{(1+e^{-a_i})^2}\right) e^{-a_i} (-\vec{x}_{ij}^T)(-1)$$

$$= \sum_{i=1}^{p} y_i \frac{1}{(1+e^{-a_i})} (e^{-a_i}) x_{ij} + \sum_{i=1}^{p} (y_i - 1) \frac{1}{1+e^{-a_i}} x_{ij}$$

$$= \sum_{i=1}^{p} (y_i - sig(a_i)) \cdot x_{ij}$$

$$\therefore \nabla L = \sum_{i=1}^{p} (y_i - sig(a_i)) \vec{x}_i$$

$$\nabla L = -\sum_{i=1}^{p} (sig(a_i) - y_i) \vec{x}_i$$

Q.E.D.

Problem 4:

Prof

$$\frac{\partial}{\partial w_j} \nabla L = \frac{\partial}{\partial w_j}\left\{\sum_{i=1}^{p} \left(\frac{1}{1+e^{-\vec{x}_i^T \vec{w}}} - y_i\right) \vec{x}_i\right\}$$

$$= -\sum_{i=1}^{p} \frac{1}{1+e^{-\vec{x}_i^T \vec{w}}} \cdot \left(\frac{e^{-\vec{x}_i^T \vec{w}}}{1+e^{-\vec{x}_i^T \vec{w}}}\right) \cdot \vec{x}_i \cdot (-x_{ij})$$

$$= -\sum_{i=1}^{p} sig(a_i) \frac{e^{-\vec{x}_i^T \vec{w}} + 1 - 1}{1+e^{-\vec{x}_i^T \vec{w}}} \vec{x}_i (-x_{ij})$$

$$\therefore \nabla^2 L = -\sum_{i=1}^{p} sig(a_i)(1 - sig(a_i)) \vec{x}_i \vec{x}_i^T$$

Q.E.D

Problem 5.

Prof: Given $Pr(\tilde{w} | \tilde{X}_p, \bar{y}_p) = Norm_{\tilde{w}_p}[\sigma^{-2} A_p^{-1} \tilde{X}_p \bar{y}_p, A_p^{-1}]$

add other Point,

$$Pr(\tilde{w} | \tilde{X}_{p+1}, \bar{y}_{p+1}) = Norm_{\tilde{w}_{p+1}}[\sigma^{-2} A_{p+1}^{-1} \tilde{X}_{p+1} \bar{y}_{p+1}, A_{p+1}^{-1}]$$

Sol: the Posterior probabity:

$$Pr(\tilde{w} | \tilde{X}_{p+1}, \bar{y}_{p+1}, \bar{X}_p, \bar{y}_p) = \frac{Pr(\bar{y}_{p+1} | \bar{X}_{p+1}, w) \cdot Pr(\tilde{w} | \bar{X}_p, y_p)}{Pr(\bar{y}_{p+1} / \bar{X}_{p+1})}$$

$$\propto Norm_{\tilde{w}}[\frac{1}{\sigma^2} A^{-1} \tilde{X}_p \bar{y}_p, A^{-1}] \cdot Norm_{y_{p+1}}[\bar{X}_{p+1} w, \sigma_{p+1}^2 I]$$

$$\propto \exp[-\frac{1}{2\sigma^2} \tilde{w}^T \tilde{X}_p \tilde{X}_p^T \tilde{w} + \frac{1}{\sigma^2} \bar{y}_p^T \tilde{X}_p^T \tilde{w} - 2\sigma_p^{-2} \tilde{w}^T \tilde{w} - 2\sigma_{p+1}^{-2} (y_{p+1} - \bar{X}_{p+1}^T \tilde{w})^2]$$

$$\propto \exp[-\frac{1}{2\sigma^2} \tilde{w}^T \tilde{X}_p \tilde{X}_p^T \tilde{w} + \frac{1}{\sigma^2} \bar{y}_p^T \tilde{X}_p^T \tilde{w} - 2\sigma_p^{-2} \tilde{w}^T \tilde{w} + \sigma_{p+1}^{-2} y_{p+1} \bar{X}_{p+1}^T \tilde{w} - \frac{1}{2\sigma_{p+1}^{2}} \tilde{w}^T \bar{X}_{p+1} \bar{X}_{p+1}^T \tilde{w}]$$

$$= Norm_{\tilde{w}_{p+1}}[\sigma^{-2} A_{p+1}^{-1} \tilde{X}_{p+1} \bar{y}_{p+1}, A_{p+1}^{-1}].$$

Q.E.D

Since $pr(w / \bar{X}, \bar{y}) \propto \exp(-\frac{1}{2\sigma^2}(y - X^T w)^T (y - X^T w)) \exp(-\frac{1}{2} w^T \Sigma_p^{-1} w)$

$$\propto \exp[\frac{1}{\sigma_n^2} y^T X^T \tilde{w} - \frac{1}{2} w^T (\frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1}) w]$$

donate: $Q \equiv \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1}$

$\Rightarrow \frac{1}{\sigma_n^2} y^T X^T = \bar{w}^T Q, \Rightarrow \bar{w} = \frac{1}{\sigma_n^2} Q^{-1} X y$

$\therefore p(w | X, y) \propto \exp[-\frac{1}{2}(w - \bar{w})^T (\frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1})(w - \bar{w})]$