CSG220-03 DATAMINING

HW01

Ex2

$$w^* = \text{argmax } p(w|y)$$

$$p(y|w) = \mathcal{N}(y;\ Xw,\ \sigma^2 I)$$

$$p(w) = \mathcal{N}(w;\ 0,\ s^2 I) \longrightarrow L2$$

$$p(w) = \frac{1}{2b} \exp - \frac{|w|}{b} \longrightarrow L1$$

Sol (a)

$$w_{MAP} = \underset{w}{\text{argmax}}\ p(w/y)$$

$$= \underset{w}{\text{argmax}} \frac{p(y|w)\, p(w)}{p(y)}$$

$$= \underset{w}{\text{argmax}}\ p(y|w)\, p(w)$$

$$\log(w_{MAP}) = \underset{w}{\text{argmax}} \left[\log(p(y|w)) + \log(p(w)) \right] \quad \text{——— (1)}$$

Now FOR L2

$$\log P(w) = \frac{-1}{2} w^T . (s^2 I)^{-1} . w + \text{constant}$$

$$= \frac{-1}{2s^2} w . w^T + \text{constant} \quad \text{——— (2)}$$

$$\log P(y|w) = \log N(\vec{y}\,;\, X\vec{w}, \sigma^2 I)$$
$$= -\frac{1}{2}(\vec{y} - X\vec{w})^T.\ (\sigma^2 I)^{-1}(\vec{y} - X\vec{w})$$
$$= -\frac{1}{2\sigma^2}(\vec{y} - X\vec{w})(\vec{y} - X\vec{w})^T \quad \text{--- (3)}$$

Combining (1), (2), (3)

$$\log w_{map} = \underset{w}{\operatorname{argmax}}\left[-\frac{1}{2\sigma^2}(X\vec{w} - \vec{y})(X\vec{w} - \vec{y})^T - \frac{1}{2S^2} w \cdot w^T\right]$$
$$= \underset{w}{\operatorname{argmax}}\left[-\frac{1}{2\sigma^2}\|Xw - Y\|^2 - \frac{1}{2S^2}\|w\|^2\right]$$
$$= \underset{N}{\operatorname{argmin}}\left[\frac{1}{N}\|Xw - y\|^2 + \frac{\sigma^2}{NS^2}\|w\|\right]^2$$

For L2, $E(w) = \frac{1}{N}\|Xw - y\|^2 + \lambda\|w\|^2$

Thus ridge regression is equivalent to MAP estimate

where $\lambda = \frac{\sigma^2}{NS^2}$

FOR L1

$$\log P(w) = \log\left(\frac{1}{2b}\exp -\frac{|w|}{b}\right)$$
$$= -\log 2b - \frac{|w|}{b}$$
$$= -\frac{|w|}{b} \quad \text{--- (4)}$$

Putting (1), (2), (4)

$$\log w_{map} = \underset{w}{\operatorname{argmax}} \left[\frac{-1}{2\sigma^2} \|Xw - Y\|^2 - \frac{\|w\|}{b} \right]$$

$$= \underset{w}{\operatorname{argmin}} \left[\frac{1}{N} \|Xw - Y\|^2 - \frac{2\sigma^2}{b} \|w\| \right]$$

& L1 (LASSO) minimizes to,

$$E(w) = \frac{1}{N} \|Xw - y\|^2 + \lambda |w|$$

Thus Lasso is equivalent to MAP estimate where $\lambda = \frac{2\sigma^2}{b}$.

Sol (b)

$$p(w) = N(w; m_0, S_0)$$
$$p(w|y, m_0, S_0) = N(w : m, S)$$
$$p(y|w) = N(y : Xw, \sigma^2 I)$$

Now if $p(x) = N(x|\mu, \Lambda^{-1})$

& $p(y|x) = N(y|Ax + b, L^{-1})$

then the marginal distribution of y and conditional distribution of x given y are

$$p(y) = N(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$$

$$\& \; p(x|y) = N(x|\Sigma\{A^T L(y-b) + \Lambda\mu\}, \Sigma)$$

$$\Sigma = (\Lambda + A^T L A)^{-1}$$

Thus $P(w|y, m_0, S_0)$

$$= N(x \mid \Sigma\{A^T L(y-b) + A\mu\}, \Sigma)$$

As $P(w)$ & $P(y|w)$ are both normal

where

$x = w$ $\quad b = 0$

$\mu = m_0$ $\quad A = X$

$\Lambda^{-1} = S_0$ $\quad L = (\sigma^2 I)^{-1}$

$\Rightarrow P(w|y, m_0, S_0)$

$$= N(w \mid \Sigma\{X^T \cdot (\sigma^2 I)^{-1}(y) + (S_0)^{-1} \cdot m_0\}, \Sigma)$$

& $\Sigma = ((S_0)^{-1} + X^T(\sigma^2 I)^{-1})^{-1}$

Hence

mean, $m = \Sigma\{X^T(\sigma^2 I)^{-1}(y) + (S_0)^{-1} m_0\}$

covariance, $S = \Sigma$

where $\Sigma = ((S_0)^{-1} + X^T(\sigma^2 I)^{-1})^{-1}$