ECE5984 Homework2

소프트웨어학과 2022710836 김민근

1. Ridge regression

(a).

$$\theta_{(MAP)} = \underset{\theta}{argmax}\, log\, p(\theta|D) = \underset{\theta}{argmax}\, log \frac{p(D|\theta)p(\theta)}{p(D)} = \underset{\theta}{argmax}\, log\, p(\theta)p(D|\theta)$$

$$= \underset{\theta}{argmax}\, log\, p(\theta) + log\, p(D|\theta)$$

In log p(θ),

$$log\, p(\theta) = logN(\theta; 0, \gamma^2 I) = log \prod_{j=1}^{D} N(0, \gamma^2)$$

$$= log \prod_{j=1}^{D} \frac{1}{\sqrt{2\pi\gamma^2}} exp\left(-\frac{\left(0-\theta_j\right)^2}{2\gamma^2}\right)$$

$$= \sum_{j=1}^{D} log \frac{1}{\sqrt{2\pi\gamma^2}} exp\left(-\frac{1}{2\gamma^2}\theta_j^2\right) \propto \frac{1}{2\gamma^2}\sum_{j=1}^{D}\theta_j^2 = -\frac{1}{2\gamma^2}\|\theta\|_2^2$$

In log p(D|θ),

$$log\, p(D|\theta) = log \prod_{i=1}^{m} N\left(\theta^T x^{(i)}, \sigma^2\right)$$

$$log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}\left(y^{(i)} - \theta^T x^{(i)}\right)^2\right)$$

$$= \sum_{i=1}^{m} log \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}\left(y^{(i)} - \theta^T x^{(i)}\right)^2\right) \propto -\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2$$

In log p(θ)+ p(D|θ) ,

$$\underset{\theta}{argmax}\, log\, p(\theta) + log\, p(D|\theta)$$

$$= \underset{\theta}{argmax} -\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2 - \frac{1}{2\gamma^2}\|\theta\|_2^2$$

$$= \underset{\theta}{argmax} -\sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2 - \lambda\|\theta\|_2^2$$

$$= \underset{\theta}{argmin} \sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2 + \lambda\|\theta\|_2^2$$

(b).

$$Let\ \ J(\theta) = \sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2 + \lambda\|\theta\|_2^2$$

Using the design matrix notation,

$$J(\theta) = (y - X\theta)^T(y - X\theta) + \lambda\theta^T\theta$$

$$\nabla_\theta J(\theta) = X^T X\theta - X^T y + \lambda\theta$$

In $\nabla_\theta J(\theta) = 0$,

$$X^T X\theta - X^T y + \lambda\theta = 0$$

$$(X^T X + \lambda I)\theta = X^T y$$

$$\theta = (X^T X + \lambda I)^{-1} X^T y$$

(c).

$$Let\ \hat{X} = \begin{pmatrix} X \\ \sqrt{\lambda} I_{n\times n} \end{pmatrix}, \hat{y} = \begin{pmatrix} y \\ 0_{n\times 1} \end{pmatrix}$$

In ordinary least squares regression,

$$\hat{\theta} = \left(\hat{X}^T\hat{X}\right)^{-1}\hat{X}^T\hat{y}$$

$$Let\ \ \hat{X}^T\hat{X} = X^T X + \lambda I,\ \hat{X}^T\hat{y} = X^T y\,,$$

$$\hat{\theta} = (X^T X + \lambda I)^{-1} X^T y$$

$\therefore$ *So*, ordinary least squares regression on an augmented data can achieve the ridge regression estimates

(d).

$$Let\ \Phi_{i,j} = \phi\left(x^{(i)}\right)^T \phi\left(x^{(j)}\right) = K_{ij}$$

$$\theta = \left(\Phi^T\Phi + \lambda I\right)^{-1}\Phi^T y$$

$$= \Phi^T\left(\Phi\Phi^T + \lambda I\right)^{-1} y \quad \cdots (\because (\lambda I + BA)^{-1}B = B(\lambda I + AB)^{-1})$$

$$= \Phi^T(\mathrm{K} + \lambda I)^{-1} y$$

$$\therefore \theta^T = y^T(\mathrm{K} + \lambda I)^{-1}\Phi$$

$$y_{new} = \theta^T \phi(x_{new})$$

$$= y^T (\mathrm{K} + \lambda I)^{-1} \Phi \phi(x_{new})$$

$$= \sum_{\mathrm{i=1}}^{\mathrm{m}} y^T (\mathrm{K} + \lambda I)^{-1} K\left(x^{(i)}, x_{new}\right)$$

$\therefore$ *we can make predictions without ever explicitly compute* $\phi\left(x^{(i)}\right)$