HW-0

S. Supralohith Reddy
EE15BTECH11026

1)

$$X = \begin{bmatrix} x_0^1 & x_1^1 & x_2^1 & \cdots & x_M^1 \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_M^2 \\ \vdots & & & & \vdots \\ x_0^N & x_1^N & x_2^N & \cdots & x_M^N \end{bmatrix} ; \quad \vec{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_M \end{bmatrix}$$

(where $x_0^{(i)} = 1 \cdot 1 \le i \le N$)

$$\vec{\hat{y}} = X \cdot \vec{w}$$

$$\Rightarrow \hat{y}^{(i)} = w_0 + \sum_{j=1}^{M} x_j^{(i)} \cdot w_j$$ (Simplest case of linear regression)

$$E(\vec{w}) = \arg\min_{\vec{w} \in R^{M+1}} \sum_{i=1}^{N} \left(y^{(i)} - \left(w_0 + \sum_{j=1}^{M} x_j^{(i)} \cdot w_j\right)\right)^2$$

Set $\frac{\partial E(\vec{w})}{\partial w_k} = 0$ for $0 \le k \le M$ and solve for $\vec{w}^*$.

$$\frac{\partial E(\vec{w})}{\partial w_k} = \frac{\partial}{\partial w_k}\left(\sum_{i=1}^{N}\left(y^{(i)} - \sum_{j=0}^{M} x_j^{(i)} \cdot w_j\right)^2\right)$$

$$= -2 \cdot \left(\sum_{i=1}^{N}\left(y^{(i)} - \sum_{j=0}^{M} x_j^{(i)} \cdot w_j\right) x_k^{(i)}\right)$$

$$= -2\left[x_k^{(1)} \cdot \left(y^{(1)} - \sum_{j=0}^{M} x_j^{(1)} \cdot w_j\right) + x_k^{(2)} \cdot \left(y^{(2)} - \sum_{j=0}^{M} x_j^{(2)} \cdot w_j\right) + \cdots\right]$$

$$= -2 \cdot \begin{bmatrix} x_k^{(1)} & \cdots & x_k^{(N)} \end{bmatrix} \cdot \begin{bmatrix} y^{(1)} - \sum_{j=0}^{M} x_j^{(1)} \cdot w_j \\ \vdots \\ y^{(N)} - \sum_{j=0}^{M} x_j^{(N)} \cdot w_j \end{bmatrix} = 0$$

Similarly,

$$\nabla E(\vec{w}) = \begin{bmatrix} \frac{\partial E(\vec{w})}{\partial w_0} \\ \vdots \\ \frac{\partial E(\vec{w})}{\partial w_M} \end{bmatrix} = -2 \begin{bmatrix} x_0^{(1)} & x_0^{(2)} & \cdots & x_0^{(N)} \\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(N)} \\ \vdots & & & \vdots \\ x_M^{(1)} & x_M^{(2)} & \cdots & x_M^{(N)} \end{bmatrix} \cdot (\vec{y} - X\vec{w})$$

$$= -2 \cdot X^T \cdot (\vec{y} - X\vec{w}) = 0$$

$$\Rightarrow X^T \vec{y} - X^T X \vec{w} = 0 \Rightarrow \boxed{\vec{w}^* = (X^T X)^{-1} \cdot X^T \vec{y}}$$

2)

$$\phi(x) = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \cdots & \phi_M(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \cdots & \phi_M(x_2) \\ \vdots & & & \\ \phi_0(x_N) & \phi_1(x_N) & \cdots & \phi_M(x_N) \end{bmatrix}$$

$$\vec{y} = \phi(x)\cdot\vec{w};$$ $\hookrightarrow \phi_0(x_i) = 1$ where $1 \le i \le N$.

$$\hat{y}^{(i)} = w_0 + \sum_{j=1}^{M} w_j \cdot \phi_j(x_i)$$

$$\Rightarrow \hat{y}^{(i)} = \sum_{j=0}^{M} w_j \cdot \phi_j(x_i)$$

$$E(\vec{w}) = \underset{\vec{w} \in R^{M+1}}{\arg\min} \sum_{i=1}^{N} \left(y^{(i)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_i)\right)^2$$

Set $\frac{\partial E(\vec{w})}{\partial w_k} = 0$ for $0 \le k \le M$ and solve for $\vec{w}^*$:

$$\frac{\partial E(\vec{w})}{\partial w_k} = -2\left(\sum_{i=1}^{N}\left(y^{(i)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_i)\right)\cdot\phi_k(x_i)\right)$$

$$= -2\left[\phi_k(x_1)\left(y^{(1)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_1)\right) + \phi_k(x_2)\left(y^{(2)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_2)\right) + \cdots\right]$$

$$= -2\begin{bmatrix} \phi_k(x_1) & \phi_k(x_2) & \cdots & \phi_k(x_N) \end{bmatrix} \begin{bmatrix} y^{(1)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_1) \\ \vdots \\ y^{(N)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_N) \end{bmatrix}$$

$$\nabla E(\vec{w}) = -2\begin{bmatrix} \phi_0(x_1) & \phi_0(x_2) & \cdots & \phi_0(x_N) \\ \phi_1(x_1) & \phi_1(x_2) & \cdots & \phi_1(x_N) \\ \vdots & & & \\ \phi_M(x_1) & \phi_M(x_2) & \cdots & \phi_M(x_N) \end{bmatrix} \begin{bmatrix} y^{(1)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_1) \\ \vdots \\ y^{(N)} - \sum_{j=0}^{M} w_j \cdot \phi_j(x_N) \end{bmatrix}$$

$$= -2 \cdot \phi(x)^T \cdot \left(\vec{y} - \phi(x)\cdot\vec{w}\right) = 0$$

$$= \phi(x)^T \cdot \phi(x) - \phi(x)^T \cdot \phi(x)\cdot\vec{w} = 0.$$

$$\Rightarrow \phi(x)^T \cdot \phi(x) - \phi(x)^T \cdot \phi(x) \cdot \vec{w}$$

Scanned by CamScanner

$$\Rightarrow \quad \boxed{\vec{w}^* = \left(\phi(x)^T \cdot \phi(x)\right)^{-1} \cdot \phi(x)^T \cdot \vec{y}}$$

3)
$$\sigma(x) = \frac{1}{1+\exp(-x)} = \frac{1}{1+e^{-x}}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1-e^{-2x}}{1+e^{-2x}},$$

$$= \frac{2-(1+e^{-2x})}{1+e^{-2x}} = \frac{2}{1+e^{-2x}} - 1$$

$$\tanh(x) = 2\sigma(2x) - 1$$

$$\hat{y}(x,w) = w_0 + \sum_{j=1}^{M} w_j \cdot \sigma\left(\frac{x-\mu_j}{s}\right)$$

Since we have $\tanh(x) = 2\sigma(2x) - 1$ which is linearly equated, we can write general linear combination of logistic sigmoid functions is equivalent to a general linear combination of 'tanh' functions.

$$\Rightarrow \sigma(2x) = \frac{1+\tanh(x)}{2}$$

$$\Rightarrow \sigma(x) = \frac{1+\tanh\left(\frac{x}{2}\right)}{2}$$

$$\hat{y}(x,w) = w_0 + \sum_{j=1}^{M} w_j \cdot \sigma\left(\frac{x-\mu_j}{s}\right)$$

$$= w_0 + \sum_{j=1}^{M} w_j \cdot \left[\frac{1+\tanh\left(\frac{x-\mu_j}{2s}\right)}{2}\right]$$

$$= w_0 + \sum_{j=1}^{M} \left(\frac{w_j}{2} + \frac{w_j}{2} \cdot \tanh\left(\frac{x-\mu_j}{2s}\right)\right)$$

$$= w_0 + \frac{w_j \cdot M}{2} + \sum_{j=1}^{M} \frac{w_j}{2} \cdot \tanh\left(\frac{x-\mu_j}{2s}\right)$$

Comparing it with $\hat{y}(x,u) = u_0 + \sum_{j=1}^{M} u_j \cdot \tanh\left(\frac{x-\mu_j}{2s}\right)$

$\therefore u_0 = w_0 + \frac{w_j \cdot M}{2}$; $u_j = \frac{w_j}{2}$ where $1 \leq j \leq M$

4) When Y is a K dimension vector

$$\vec{y} = \begin{bmatrix} y_1^{(1)} & y_2^{(1)} & \cdots & y_K^{(1)} \\ \vdots & & & \\ y_1^{(N)} & y_2^{(N)} & \cdots & y_K^{(N)} \end{bmatrix}_{N \times K}$$

And for each dimension of y, there will be different weights. $W \rightarrow (M+1) \times K$

$$\vec{w} = \begin{bmatrix} w_0^{(1)} & w_0^{(2)} & \cdots & w_0^{(K)} \\ \vdots & & & \\ w_M^{(1)} & w_M^{(2)} & \cdots & w_M^{(K)} \end{bmatrix}_{(M+1) \times K}$$

$\vec{y} = X . \vec{w}$ (simplest case of linear regression)

$$E(\vec{w}) = \arg\min_{\vec{w}} \sum_{t=1}^{K} \sum_{i=1}^{N} \left(y_t^{(i)} - \sum_{j=0}^{M} x_j^{(i)} . w_j^t \right)^2$$

$$\frac{\partial E(\vec{w})}{\partial w_m^{(l)}} = \arg\min_{\vec{w}} \cancel{\sum_{t=1}^{K}} \sum_{i=1}^{N} \left(-2 . \left(y_t^{(i)} - \sum_{j=0}^{M} x_j^{(i)} . w_j^t \right) . x_m^{(i)} \right)$$

$$= -2 \left[x_m^{(1)} \left[\cancel{\sum_{t=1}} \left(y_t^{(1)} - \sum_{j=0} x_j^{(1)} . w_j^{(l)} \right) \right] + x_m^{(2)} \left(\cancel{\sum_{t=1}} y_t^{(2)} - \sum_{j=0} x_j^{(1)} . w_j^{(l)} \right) + \cdots \right]$$

$$= -2 . \left[x_m^{(1)} \; x_m^{(2)} \cdots x_m^{(N)} \right] \begin{bmatrix} \cancel{\sum_{t=1}} \left(y_t^{(1)} - \sum_{j=0}^{M} x_j^{(1)} . w_j^{(l)} \right) \\ \vdots \\ \cancel{\sum_{t=1}} \left(y_t^{(N)} - \sum_{j=0}^{M} x_j^{(N)} . w_j^{(l)} \right) \end{bmatrix}$$

$\cancel{H_t \sum_{t=1}^{K} \vec{y}_t^{(i)} (y}$

$$\hat{y}_l^{(i)} = \sum_{j=0}^{M} x_j^{(i)} . w_j^{(l)}$$

Scanned by CamScanner

$$\Rightarrow \begin{bmatrix} \frac{\partial E(\vec{w})}{\partial w_0^{(1)}} & \frac{\partial E(\vec{w})}{\partial w_0^{(2)}} & \cdots & \frac{\partial E(\vec{w})}{\partial w_0^{(K)}} \\ \vdots & & & \\ \frac{\partial E(\vec{w})}{\partial w_M^{(1)}} & \frac{\partial E(\vec{w})}{\partial w_M^{(2)}} & \cdots & \frac{\partial E(\vec{w})}{\partial w_M^{(K)}} \end{bmatrix}$$

$$= -2 \begin{bmatrix} x_0^{(1)} & x_0^{(2)} & \cdots & x_0^{(N)} \\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(N)} \\ \vdots & & & \\ x_M^{(1)} & x_M^{(2)} & \cdots & x_M^{(N)} \end{bmatrix} \begin{bmatrix} y_1^{(1)} - \sum_{j=0}^{M} x_j^{(1)} w_j^{(1)} & \cdots & y_K^{(1)} - \sum x_j^{(1)} w_j^{(K)} \\ \vdots & & \\ y_1^{(N)} - \sum_{j=0}^{M} x_j^{(N)} \cdot w_j^{(1)} & \cdots & y_K^{(N)} - \sum x_j^{(N)} w_j^{K} \end{bmatrix}$$

$$= -2 \cdot X^T \cdot (Y - XW) = 0$$

$$\Rightarrow X^T Y - X^T X W = 0 \Rightarrow W = (X^T X)^{-1} X^T Y$$

$$\therefore \boxed{\vec{w}^* = (X^T X)^{-1} X^T \vec{Y}}$$

Similarly, by replacing $x_j^{(i)}$ with $\phi_j(x_i)$, we can show with basis functions.

Since, it is linear with respect to $\vec{w}$ and not x.

In case of basis function $\phi(x)$,

$$\vec{w}^* = \left(\phi(x)^T \phi(x)\right)^{-1} \phi(x)^T \vec{Y}$$

Scanned by CamScanner

5) $$E(\vec{w}) = \sum_{i=1}^{N} r_i \cdot \left(y^{(i)} - \sum_{j=0}^{d} x_j^{(i)} \cdot w_j\right)^2$$

Set $\dfrac{\partial E(\vec{w})}{\partial w_k} = 0$ for $0 \le k \le d$

$$\frac{\partial E(\vec{w})}{\partial w_k} = \sum_{i=1}^{N} r_i \cdot (-2)\left(y^{(i)} - \sum_{j=0}^{d} x_j^{(i)} \cdot w_j\right) \cdot x_k^{(i)}$$

$$= -2 \sum_{i=1}^{N} r_i \cdot x_k^{(i)} \cdot \left[y^{(i)} - \sum_{j=0}^{d} x_j^{(i)} \cdot w_j\right]$$

$$= -2 \begin{bmatrix} r_1 x_k^{(1)} & \cdots\cdots & r_N x_k^{(N)} \end{bmatrix} \begin{bmatrix} y^{(1)} - \sum_{j=0}^{d} x_j^{(1)} \cdot w_j \\ \vdots \\ y^{(N)} - \sum_{j=0}^{d} x_j^{(N)} \cdot w_j \end{bmatrix}$$

Similarly,

$$\begin{bmatrix} \frac{\partial E(\vec{w})}{\partial w_0} \\ \vdots \\ \vdots \\ \frac{\partial E(\vec{w})}{\partial w_d} \end{bmatrix} = -2 \underbrace{\begin{bmatrix} r_1 x_0^{(1)} & \cdots\cdots & r_N x_0^{(N)} \\ \vdots & & \\ \vdots & & \\ r_1 x_k^{(1)} & \cdots\cdots & r_N x_k^{(N)} \end{bmatrix}}_{X_R^T \to \text{ weighted } x \text{ transpose}} \begin{bmatrix} y^{(1)} - \sum_{j=0}^{d} x_j^{(1)} \cdot w_j \\ \vdots \\ \vdots \\ y^{(N)} - \sum_{j=0}^{d} x_j^{(N)} \cdot w_j \end{bmatrix}$$

$$= X_R^T (Y - XW) * (-2) = 0$$

$$\Rightarrow X_R^T \cdot Y - X_R^T \cdot XW = 0$$

$$\Rightarrow \boxed{\vec{w}^* = (X_R^T X)^{-1} X_R^T Y}$$

Scanned by CamScanner

6) $$E(\vec{w}) = \arg\min_{\vec{w}} \left(\sum_{i=1}^{N} \left(y^{(i)} - \sum_{j=0}^{d} x_j^{(i)} \cdot w_j\right)^2 \right) + \sum_{j=0}^{d} w_j^2 \cdot \lambda$$

Set $\frac{\partial E(\vec{w})}{\partial w_k} = 0$ for $0 \le k \le d$

$$\frac{\partial E(\vec{w})}{\partial w_k} = \left(\sum_{i=1}^{N} -2 \cdot x_k^{(i)} \left(y^{(i)} - \sum_{j=0}^{d} x_j^{(i)} \cdot w_j\right)\right) + 2\lambda w_k$$

$$= -2 \cdot \begin{bmatrix} x_k^{(1)} & \cdots & x_k^{(N)} \end{bmatrix} \begin{bmatrix} y^{(1)} - \sum_{j=0}^{d} x_j^{(1)} w_j \\ \vdots \\ y^{(N)} - \sum_{j=0}^{d} x_j^{(N)} w_j \end{bmatrix} + 2\lambda \begin{bmatrix} w_0 & \cdots & w_d \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{d+1 \times 1}$$

Similarly,

$$\begin{bmatrix} \frac{\partial E(\vec{w})}{\partial w_0} \\ \vdots \\ \frac{\partial E(\vec{w})}{\partial w_d} \end{bmatrix} = -2 \begin{bmatrix} x_0^{(1)} & \cdots & x_0^{(N)} \\ \vdots & & \\ x_d^{(N)} & \cdots & x_d^{(N)} \end{bmatrix} \begin{bmatrix} Y - XW \end{bmatrix} + 2\lambda \begin{bmatrix} w_0 & \cdots & w_d \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \cdot & & 1 \end{bmatrix}$$

$$= -2 \cdot X^T \cdot (Y - XW) + 2\lambda W I = 0$$

$$\Rightarrow -X^T Y + X^T X W + \lambda W I = 0.$$

$$\Rightarrow (X^T X + \lambda I) W = + X^T Y$$

$$\Rightarrow \cancel{\vec{w}^* = (X^T X)^{-1}}$$

$$\Rightarrow \boxed{\vec{w}^* = (X^T X + \lambda I)^{-1} X^T Y}$$

The model will have a low accuracy if it is overfitting. This happens because model is trying hard to capture noise in training dataset because of which we get wrong predictions.

7) By noise, we mean the data points that don't really represent the true properties of our data. So, to avoid overfitting, we use the regularisation technique.

---

7) $\hat{y}_n \rightarrow$ calculated output in case of noise to input

$y_n \rightarrow$ calculated output without noise

$t_n \rightarrow$ expected actual output.

$$\cancel{\hat{y}_n = w_0 + \sum_{j=1}^{d}}$$

$$\cancel{\hat{y}_n = w_0 + \sum_{j=1}^{d} (x_j^{(n)} + \epsilon_j^{(n)})}$$

$$\hat{y}_n = w_0 + \sum_{j=1}^{d} (x_j^{(n)} + \eta_n) . w_j$$

$$\hat{y}_n = y_n + \sum_{j=1}^{d} \eta_n . w_j \quad \text{--- (1)}$$

$$\tilde{E}(\vec{w}) = \sum_{n=1}^{N} (\hat{y}_n - t_n)^2$$

$$= \sum_{n=1}^{N} (\hat{y}_n^2 - 2\hat{y}_n t_n + t_n^2) \quad \text{--- (2)}$$

Substituting (1) in (2)

$$\tilde{E}(\vec{w}) = \sum_{n=1}^{N} \Big(y_n^2 + 2y_n \sum_{j=1}^{d} \eta_n . w_j + \Big(\sum_{j=1}^{d} w_j . \eta_n\Big)^2 - 2t_n y_n - 2t_n \sum_{j=1}^{d} w_j . \eta_n + t_n^2 \Big)$$

$$\Rightarrow \tilde{E}(\vec{w}) = \sum_{n=1}^{N} \Big[(y_n - t_n)^2 + \Big(\sum_{j=1}^{d} w_j . \eta_n\Big)^2 + 2y_n . \sum_{j=1}^{d} \eta_n . w_j - 2t_n . \sum_{j=1}^{d} w_j \eta_n\Big]$$

We know that $E[\eta_i] = 0$ and $E[\eta_i . \eta_j] = \delta_{ij} . \sigma^2$

$$\Rightarrow E[\eta_i^2] = \sigma^2$$

Taking expectation on both sides.

We get 3rd & 4th term $= 0$ because $E[\eta_i] = 0$.

$$E\left[\sum_{n=1}^{N}\left(\sum_{j=1}^{d} w_j \eta_n\right)^2\right] = \sum_{j=1}^{d} w_j^2 \cdot \sigma^2 \quad \left(\text{since } E[\eta_i^2] = \sigma^2,\ E[\eta_i \cdot \eta_j] = 0 \text{ if } i \neq j\right)$$

$$\Rightarrow E\left[\tilde{E}(\vec{w})\right] = \sum_{n=1}^{N}(y_n - t_n)^2 + \underbrace{\sum_{j=1}^{d} w_j^2 \sigma^2}_{\text{weight regularisation term without } w_0}$$

Hence, proved

---

8) ~~$p(\vec{w}\,|\,x, \vec{y}, \alpha, \beta) = p(\vec{y}\,|\,x, \vec{w}, \beta)$.~~

8)
$$p(\vec{w}\,|\,x, \vec{y}, \alpha, \sigma) = \frac{p(\vec{y}\,|\,x, \vec{w}, \sigma) \cdot p(\vec{w}\,|\,\alpha)}{\int p(y|x, w) \cdot p(w)\, dw}$$

The denominator term is independent of w.

So, $$p(\vec{w}\,|\,x, \vec{y}, \alpha, \sigma) = \frac{p(\vec{y}\,|\,x, \vec{w}, \sigma) \cdot p(\vec{w}\,|\,\alpha)}{K}$$

$$\Rightarrow \log(p(\vec{w}\,|\,x, \vec{y}, \alpha, \sigma)) = -\log K + \log\left[(p(\vec{y}\,|\,x, \vec{w}, \sigma)) \cdot p(\vec{w}\,|\,\alpha)\right]$$

$$\Rightarrow \log(p(\vec{w}\,|\,x, \vec{y}, \alpha, \sigma)) = -\log K + \log\left[\frac{1}{(2\pi)^{N/2} (\sigma^2)^{N/2}} \cdot e^{-\frac{1}{2}(\vec{y} - x\vec{w})^T (\sigma^2 I)^{-1} (\vec{y} - x\vec{w})} \cdot \frac{1}{(2\pi)^{N/2} (\alpha^2)^{N/2}} e^{-\frac{1}{2}\vec{w}^T(\alpha^2 I)\vec{w}}\right]$$

$$\Rightarrow \log(p(\vec{w}\,|\,x, \vec{y}, \alpha, \sigma)) = k' - \frac{1}{2\sigma^2}(\vec{y} - x\vec{w})^T(\vec{y} - x\vec{w}) - \frac{1}{2\alpha^2}\vec{w}^T\vec{w} \quad (1)$$

$$\Rightarrow \log(p(\vec{w}\,|\,x, \vec{y}, \alpha, \sigma)) = \frac{1}{2\sigma^2}\left(k'' - \left[(\vec{y} - x\vec{w})^T(\vec{y} - x\vec{w}) + \frac{\sigma^2}{\alpha^2}\vec{w}^T\vec{w}\right]\right)$$

We can observe that maximizing p will give same result as ridge regression with $\lambda = \frac{\sigma^2}{\alpha^2}$

$$\therefore \vec{w}^* = \left(X^T X + \frac{\sigma^2}{\alpha^2} I\right)^{-1} X^T \vec{y}$$

Now, we need to show that ridge regression estimate is the mean of posterior distribution:

Assume $p(\vec{w} \mid \vec{x}, \vec{y}, \alpha, \sigma)$ to be gaussian with mean 'm' and variance V.

The exponent term will be $= \frac{-1}{2}(w-m)^T V^{-1}(w-m)$

On comparing with ① → (On comparing with ②)

2^{nd} power of w comparison ⇒

$$\frac{-1}{2}\vec{w}^T V^{-1}\vec{w} = \frac{-1}{2\sigma^2}\vec{w}^T X^T X\vec{w} - \frac{1}{2\alpha^2}\vec{w}^T\vec{w}$$

$$\therefore V^{-1} = \frac{1}{\sigma^2} X^T X + \frac{I}{\alpha^2}$$

1^{st} power of w comparison →

$$\frac{1}{2}\vec{w}^T V^{-1} m = \frac{1}{2\sigma^2}\vec{w}^T X^T \vec{y}$$

$$\Rightarrow V^{-1} m = \frac{1}{\sigma^2} X^T \vec{y}$$

$$\Rightarrow m = \frac{1}{\sigma^2} V X^T \vec{y}$$

$$\therefore m = \frac{1}{\sigma^2}\left[\frac{1}{\sigma^2} X^T X + \frac{I}{\alpha^2}\right]^{-1} X^T \vec{y}$$

$$= \left[X^T X + \frac{\sigma^2}{\alpha^2} I\right]^{-1} X^T \vec{y}$$

$\therefore$ The mean of posterior distribution of $\vec{w}$ is exactly same as ridge regression estimate with $\lambda = \frac{\sigma^2}{\alpha^2}$.