Homework #3.

1. (a). – $\mu = 0.01$

$tr(R) = 10, \quad d = 3$

$$\tau_{av} \approx \frac{d+1}{4\mu \cdot tr(R)} = \frac{4}{4 \times 0.01 \times 10} = 10$$

– $\mu = 0.1$

$$\tau_{av} \approx \frac{4}{4 \times 0.1 \times 10} = 1.$$

(b) $\mu = 0.01$

$$M \approx \frac{d+1}{4\tau_{av}} = \frac{4}{4 \times 10} = 0.1$$

$\mu = 0.1$

$$M \approx \frac{d+1}{4\tau_{av}} = \frac{4}{4 \cdot 1} = 1$$

2. (a). $\nabla_k = 2R\vec{w}_k - 2\vec{p}$

$$\therefore \vec{w}_{k+1} = \vec{w}_k - \mu R^{-1} \nabla_k$$
$$= \vec{w}_k - \mu R^{-1}(2R\vec{w}_k - 2\vec{p})$$
$$= \vec{w}_k - 2\mu\vec{w}_k + 2\mu R^{-1}\vec{p}$$
$$= (1-2\mu)\vec{w}_k + 2\mu \vec{w}^*.$$

Take $\mu = \frac{1}{2}$,

then $\vec{w}_1 = (1 - 2 \times \frac{1}{2})\vec{w}_0 + 2 \times \frac{1}{2}\vec{w}^*$

$$\therefore \vec{w}_1 = \vec{w}_*$$

i.e. we need just one step if $\mu = \frac{1}{2}$.

(b) If R is singular $\Rightarrow$ using $R = Q\Lambda Q^{-1}$

where $Q = \begin{bmatrix} \cdots & v_i & \cdots \end{bmatrix}$, v_i = eigen vectors of R

$\Lambda = \begin{bmatrix} \ddots & & 0 \\ & \lambda_i & \\ 0 & & \ddots \end{bmatrix}$, λ_i = eigen values of R.

Since R = symmetric, Q = orthonormal

Since R = singular, $\Lambda = \left[\begin{array}{c|c} \lambda_1 \ddots \lambda_k & 0 \\ \hline 0 & 0 \end{array}\right]$

Since R^{-1} does not exists, consider the pseudo-inverse R' of R.

$$R' = Q\Lambda' Q^T = \left[\begin{array}{c|c} \Lambda_k^{-1} & 0 \\ \hline 0 & 0 \end{array}\right]$$

Take $\omega' = R'\omega$ and check the MSE.

$$\|R\omega' - P\|^2 = \|Q\Lambda Q^T(Q\Lambda' Q^T) - P\|^2$$

$$= \|Q\Lambda\Lambda' Q^T - P\|^2$$

$$= \|I_k Q^T - Q^T P\|^2$$

= smallest possible value where Q & P are given and $\nexists R^{-1}$

3. <1st iteration>

$k=0$, $\vec{\omega}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $S_0 = I$.

(step 1) $\nabla E = \begin{bmatrix} +1 + 4\omega_1 + 2\omega_2 \\ -1 + 2\omega_1 + 2\omega_2 \end{bmatrix}$

$\Rightarrow \nabla_0 = \nabla E \big|_{\vec{\omega} = [\,^0_0\,]} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$\therefore \vec{d_0} = -S_0 \cdot \nabla_0 = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

(step 2) $E(\vec{\omega}_0 + \alpha_0 \vec{d_0}) = E\left(\begin{bmatrix} -\alpha_0 \\ \alpha_0 \end{bmatrix}\right)$

$= -\alpha_0 - \alpha_0 + 2\alpha_0^2 - 2\alpha_0^2 + \alpha_0^2$

$= \alpha_0^2 - 2\alpha_0 = (\alpha_0 - 1)^2 - 1$

$\therefore \alpha_0 = 1$

(Step 3) $\vec{w}_1 = \vec{w}_0 + \alpha_0 \vec{d}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\vec{p}_0 = \alpha_0 \vec{d}_0 = 1 \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\vec{q}_0 = \nabla_1 - \nabla_0 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$

(Step 4) $S_1 = S_0 + \dfrac{\vec{p}_0 \vec{p}_0^{\,T}}{\vec{p}_0^{\,T} \vec{q}_0} - \dfrac{S_0 \vec{q}_0 \vec{q}_0^{\,T} S_0}{\vec{q}_0^{\,T} S_0 \vec{q}_0} = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}$

(Step 5) $E(\vec{w}_1) = E(-1, 1) = -1$

<2nd iteration>

(Step 1) $\nabla_1 = \nabla E \big|_{\vec{w} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$\therefore \vec{d}_1 = -S_1 \cdot \nabla_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(Step 2) $E(\vec{w}_1 + \alpha_1 \vec{d}_1) = E\left(\begin{bmatrix} -1+0 \\ 1+\alpha_1 \end{bmatrix}\right)$

$= \alpha_1^2 - \alpha_1 - 1 = (\alpha_1 - \frac{1}{2})^2 - \frac{5}{4}$

$\therefore \alpha_1 = \frac{1}{2}$

(Step 3) $\vec{w}_2 = \vec{w}_1 + \alpha_1 \vec{d}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix}$

$\vec{p}_1 = \alpha_1 \vec{d}_1 = \frac{1}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}$

$\vec{q}_1 = \nabla_2 - \nabla_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(Step 4) $S_2 = S_1 + \dfrac{\vec{p}_1 \cdot \vec{p}_1^{\,T}}{\vec{p}_1 + \vec{q}_1} - \dfrac{S_1 \vec{q}_1 \vec{q}_1^{\,T} S_1}{\vec{q}_1^{\,T} S_1 \vec{q}_1} = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1 \end{bmatrix}$

(Step 5) $E(\vec{w}_2) = E(-1, \frac{3}{2}) = -1.25$

<3rd iteration> $k = 2$

(Step 1) $\nabla_2 = \nabla E \big|_{\vec{w} = \begin{bmatrix} -1 \\ 3/2 \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\therefore \vec{w}^* = \vec{w}_2 = \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix}$

4. CD algorithm is used for solving $Q\vec{x} = \vec{b}$

(if Q is symmetric & positive-definite)

By conjugate Direction Theorem, we need only n iterations for $\{\vec{x}_n\}$ to converge to $\vec{x}^*$.

Our task is to make a linear regression model and eventually this is to solve $R\vec{w}^* = \vec{p}$

Here, R is symmetric and Positive-definite.

If $E[x_{ik}^2] \neq 0$ for $\forall i = 0, 1, \dots, d$ where x_{ik} is ith element of training vector $\vec{x}_k$

So the CD algorithm can find the optimal solution $\vec{w}^*$ after n iterations.