<Proper step size in Linear Regression>

In L.S) $(N \to \infty)$

$$\mathbf{w} = (X^T X)^{-1} X^T \mathbf{y}$$

$$\frac{1}{N} X^T X = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T = \mathbb{E}[\mathbf{x}\mathbf{x}^T] \triangleq R_x .$$

$$\frac{1}{N} X^T \mathbf{y} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i y_i = \mathbb{E}[\mathbf{x} y] \triangleq \mathbf{r}_{xy} .$$

$$\mathbf{w}^* = \left(\frac{1}{N} X^T X\right)^{-1} \left(\frac{1}{N} X^T \mathbf{y}\right)$$

$$= R_x^{-1} \mathbf{r}_{xy} .$$

In Gradient descent.) $(N \to \infty)$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \alpha\, \mathbb{E}\left[(y - \mathbf{w}(n)^T \mathbf{x})\, \mathbf{x}\right] \quad \cdots (1)$$

$$\mathbb{E}\left[(y - \mathbf{w}(n)^T \mathbf{x})\, \mathbf{x}\right] = \mathbb{E}[\mathbf{x} y] - \mathbb{E}[\mathbf{x}\mathbf{x}^T]\mathbf{w}(n)$$

$$= \mathbf{r}_{xy} - R_x \mathbf{w}(n) \quad \cdots (2)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \alpha\left(\mathbf{r}_{xy} - R_x \mathbf{w}(n)\right) \quad (\because (1), (2))$$

$$= (I - \alpha R_x)\mathbf{w}(n) + \alpha\, \mathbf{r}_{xy}$$

$$w(n+1) - w^* = (I - \alpha R_x)w(n) + \alpha \underline{r_{xy}} - w^*$$

$$= (I - \alpha R_x)w(n) + \alpha \underline{R_x w^*} - w^*$$

$$= (I - \alpha R_x)(w(n) - w^*)$$

Let $c(n+1) = (1 + \alpha R_x)c(n)$

where $c(n) \triangleq w(n) - w^*$.

$$c(n+1) = (I - \alpha U \Lambda U^T)c(n) \quad (\because \text{EVD})$$

$$= (UU^T - \alpha U \Lambda U^T)c(n)$$

$$= U(I - \alpha \Lambda)\underline{U^T c(n)} \quad \cdots (3)$$

Let $v(n+1) = (I - \alpha\Lambda)v(n)$

where $v(n) \triangleq U^T c(n)$

$$v(n+1) = (I - \alpha\Lambda)v(n)$$

$$= (I - \alpha\Lambda)(I - \alpha\Lambda)v(n-1)$$

$$\vdots$$

$$= (I - \alpha\Lambda)^n v(0)$$

$$\downarrow$$

$$v_i(n+1) = (1 - \alpha\lambda_i)^n v_i(0), \quad i = 1, 2, \cdots, m.$$

$$\downarrow$$

$$\forall i, \quad |1 - \alpha\lambda_i| < 1$$

Therefore, proper step size α is in

$\forall_i,\ |1-\alpha\lambda_i|<1.$

$$\Leftrightarrow \quad 0<\alpha<\frac{2}{\max_i \lambda_i} \qquad \blacksquare$$