Machine Learning - Assignment 3

Sudip Sinha

May 2, 2014

Exercise 2

Given

- $VV^T = V^TV = I$
- D is an invertible diagonal matrix

Proof

Let $B = VD^{-1}V^T$

- $AB = VDV^TVD^{-1}V^T = VD\left(V^TV\right)D^{-1}V^T = VDID^{-1}V^T = V\left(DD^{-1}\right)V^T = VIV^T = I$
- $BA = VD^{-1}V^TVDV^T = VD^{-1}\left(V^TV\right)DV^T = VD^{-1}IDV^T = V\left(D^{-1}D\right)V^T = VIV^T = I$

Thus, $A^{-1} = B = VD^{-1}V^T$.

Exercise 3

We just have to show that the square of the norm is a convex function.

By the definition of norm,

$$\begin{aligned} \|tx + (1-t)\,y\| &\leq \|tx\| + \|(1-t)\,y\| = t\,\|x\| + (1-t)\,\|y\| \\ \Longrightarrow \|tx + (1-t)\,y\|^2 &\leq (t\,\|x\| + (1-t)\,\|y\|)^2 \leq t^2\,\|x\|^2 + (1-t)^2\,\|y\|^2 \end{aligned}$$

Thus, the least squares loss function $\|y - Xw\|^2$ is a convex function of w.

Exercise 4

Let $W = \operatorname{diag}(r_i)$. Therefore, $W^T = W$

Thus, the problem can be reformulated as

$$w = \arg\min_w (e)\,, \text{ where } e = (y - Xw)^T W (y - Xw)$$

Using the rules of Matrix Calculus given here:

$$\begin{aligned} \frac{\partial e}{\partial w} &= (y - Xw)^T W \frac{\partial}{\partial w}(y - Xw) + (y - Xw)^T W^T \frac{\partial}{\partial w}(y - Xw) \\ &= 2\,(y - Xw)^T W^T (-X) \qquad \left(\text{Since } W^T = W\right) \\ &= 2\left(X^T W (Xw - y)\right)^T \end{aligned}$$

For minimum error:

$$\begin{aligned} & \tfrac{\partial e}{\partial w} && = 0^T \\ \Longrightarrow\ & 2\left(X^T W (Xw - y)\right)^T && = 0^T \\ \Longrightarrow\ & X^T W (Xw - y) && = 0 \\ \Longrightarrow\ & X^T W X w && = X^T y \\ \Longrightarrow\ & w && = \left(X^T W X\right)^{-1} X^T y \end{aligned}$$