HW Ch.6

Solution

Prob. 1 (10) a) $\mathbb{E}\,\|x-x^*\|^2 = \mathbb{E}(x-\vec{y}^T\vec{\theta}) = \mathbb{E}[x^2 - 2\vec{\theta}^T\vec{y}x + \vec{\theta}^T\vec{y}\vec{y}^T\vec{\theta}]$

$= \mathbb{E}[x^2] - 2\vec{\theta}^T\,\mathbb{E}[\vec{y}x] + \vec{\theta}^T\mathbb{E}[\vec{y}\vec{y}^T]\vec{\theta} = \sigma^2 - 2\vec{\theta}^T\vec{r}_{yx} + \vec{\theta}^T C_y\vec{\theta}$.

(10) b) $\frac{\partial}{\partial\vec{\theta}}\mathbb{E}\,\|x-x^*\|^2 = \frac{\partial}{\partial\vec{\theta}}(\sigma^2 - 2\vec{\theta}^T\vec{r}_{yx} + \vec{\theta}^T C_y\vec{\theta}) = 2C_y\vec{\theta} - 2\vec{r}_{yx} = 0 \quad \therefore \vec{\theta}^* = C_y^{-1}\vec{r}_{y,x}$.

(10) c) $C_y = \mathbb{E}\,\vec{y}\vec{y}^T = \mathbb{E}[(\vec{a}x+\vec{n})(\vec{a}x+\vec{n})^T] = \mathbb{E}[x^2\vec{a}\vec{a}^T + x\vec{a}\vec{n}^T + x\vec{n}\vec{a}^T + \vec{n}\vec{n}^T]$

$= \mathbb{E}[x^2\vec{a}\vec{a}^T] + \mathbb{E}[\vec{n}\vec{n}^T]$, by $\mathbb{E}[x\vec{a}\vec{n}^T] = \vec{a}\,\mathbb{E}[x]\,\mathbb{E}[\vec{n}^T] = 0$

$= \sigma^2\vec{a}\vec{a}^T + C_n$. $\quad \mathbb{E}[x\vec{n}\vec{a}^T] = \mathbb{E}[x]\,\mathbb{E}[\vec{n}]\,\vec{a}^T = 0$.

(10) d) $\vec{r}_{y,x} = \mathbb{E}[\vec{y}x] = \mathbb{E}[(\vec{a}x+\vec{n})x]$ by $x \perp\!\!\!\perp \vec{n}$ $\quad \because x \perp\!\!\!\perp \vec{n}$.

$= \mathbb{E}[\vec{a}x^2] + \mathbb{E}[\vec{n}x] = \vec{a}\,\mathbb{E}[x^2] + \mathbb{E}[\vec{n}]\,\mathbb{E}[x] = \sigma^2\vec{a}$.

(10) e) $x^* = \vec{y}^T\vec{\theta}^* = \vec{y}^T C_{yx}^{-1}\vec{r}_{yx} = \vec{y}^T\underbrace{(\sigma^2\vec{a}\vec{a}^T + C_n)}_{>0}{}^{-1}\sigma^2\vec{a}$

>0, $\because \vec{a}\vec{a}^T \geq 0$ & $C_n > 0$.

Prob. 2. (10) a) Observe first that $\mathbb{E}(x_n - f(\vec{y}_n, \vec{\theta}^*)) \leq \mathbb{E}(x_n - f(\vec{y}_n, \text{any } \vec{\theta}))^2$ — (*)

Observe second that $\mathbb{E}[MSE_2(\theta)] = \mathbb{E}[\frac{1}{N}\sum_{n\in S_2}(x_n - f(\vec{y}_n,\vec{\theta}))^2] = \frac{1}{N}\sum_{n\in S_2}\mathbb{E}(x_n - f(\vec{y}_n,\vec{\theta}))^2$ — (**)

(**) w/ $\vec{\theta}^*$ $\leq$ (**) w/ $\hat{\vec{\theta}}$ by using (*).

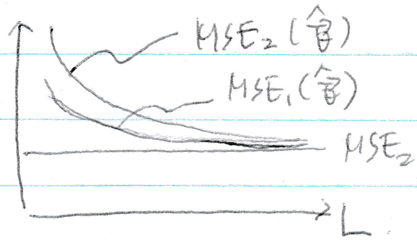
$\therefore$ $MSE_2(\vec{\theta}^*)$ is expected to be smaller than $MSE_2(\hat{\vec{\theta}})$.

$\therefore$ $MSE_2(\vec{\theta}^*)$ is expected to be smaller than $MSE_2(\hat{\vec{\theta}})$.

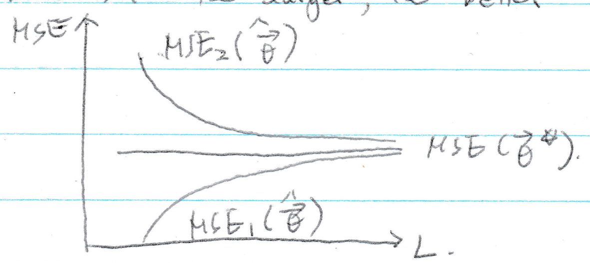
(10) b) To use $\vec{\theta}^*$, we have to know the distribution of x_n & $\vec{y}_n$.

(10) c) L should be at least larger than P; but the larger, the better $\hat{\vec{\theta}}$.

(10) d)



or



* test loss is worse than training loss, and both test and training loss approaches to the solution given by known distribution of x_n & $\vec{y}_n$.

(10) Prob. 3. 10 pts for each subproblem.