Introduction to Statistical Machine Learning
Homework 1

Yota Toyama

October 20, 2016

1.

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{w}} R(\mathbf{w}) &= 0 \\
\frac{\partial}{\partial \mathbf{w}} E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x})^2 \right] &= 0 \\
E_{p(\mathbf{x},y)} \left[2(y - \mathbf{w}^T \mathbf{x})(-\mathbf{x}) \right] &= 0 \\
E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x})\mathbf{x} \right] &= 0 \\
\mathbf{a}^T E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x})\mathbf{x} \right] &= 0 \\
E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x})\mathbf{a}^T \mathbf{x} \right] &= 0
\end{aligned}
$$

2.

$$
\begin{aligned}
E_{p(\mathbf{x},y)} \left[y - \mathbf{w}^T \mathbf{x} \right] &= 0 \\
E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x}) E_{p(\mathbf{x})} \left[\mathbf{x} \right] \right] &= 0 \\
E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x})\mathbf{x} \right] + E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x}) E_{p(\mathbf{x})} \left[\mathbf{x} \right] \right] &= 0 \\
E_{p(\mathbf{x},y)} \left[(y - \mathbf{w}^T \mathbf{x})(\mathbf{x} - E_{p(\mathbf{x})} \left[\mathbf{x} \right]) \right] &= 0
\end{aligned}
$$

$\therefore$ the corretation between data and prediction errors is 0.

3.

$$
\begin{aligned}
\hat{\mathbf{w}} &= \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \\
&= (X^T X)^{-1} X^T y \\
\text{Let } C &\in \mathbb{R}^{d+1 \times d+1} \text{ s.t. } \tilde{X} = XC \\
\hat{\tilde{\mathbf{w}}} &= \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \mathbf{w}^T \tilde{\mathbf{x}}_i)^2 \\
&= (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y \\
&= ((XC)^T XC)^{-1} (XC)^T y \\
&= (CX^T XC)^{-1} CX^T y \\
&= C^{-1} (X^T X)^{-1} C^{-1} CX^T y \\
&= C^{-1} (X^T X)^{-1} X^T y \\
\tilde{X} \hat{\tilde{\mathbf{w}}} &= XCC^{-1} (X^T X)^{-1} X^T y \\
&= X (X^T X)^{-1} X^T y \\
&= X \hat{\mathbf{w}}
\end{aligned}
$$