Homework 10 - December 8, 2016. Questions.

Intermediate Statistics (Carnegie Mellon University)

Homework 10
Due Thursday December 8 by 3:00

1. Let (X, Y) have distribution P. Let
$$R(\beta) = \mathbb{E}[(Y - \beta^T X)^2].$$
Show that the β that minimizes $R(\beta)$ is $\beta = \Lambda^{-1}\alpha$ where $\Lambda = \mathbb{E}[XX^T]$, $\alpha = (\alpha(1), \ldots, \alpha(d))$ and $\alpha(j) = \mathbb{E}[YX(j)]$.

2. Let (X, Y) have joint density $p(x, y)$. Suppose we want to predict Y from X with risk
$$R(g) = \mathbb{E}[\ |Y - g(X)|\ ].$$
Let $m(x)$ be the median of $p(y|x)$. (Assume that $m(x)$ is unique.) Show that, for any g,
$$R(g) \geq R(m).$$

3. Consider regression data $(X_1, Y_1), \ldots, (X_n, Y_n)$ where
$$Y_i = \beta X_i + \epsilon_i$$
where $X_i \in \mathbb{R}$ and $\epsilon_i \sim N(0, \sigma^2)$. Unfortunately, we do not get to observe X_i directly. Instead, we observe a noisy version of X_i. Thus, the observed data are $(W_1, Y_1), \ldots, (W_n, Y_n)$ where
$$W_i = X_i + \delta_i$$
and $\delta_i \sim N(0, \tau^2)$. We will estimate β by using the least squares estimator based on the observed data:
$$\widehat{\beta} = \frac{\sum_{i=1}^n Y_i W_i}{\sum_{i=1}^n W_i^2}.$$
Prove that $\widehat{\beta}$ is not consistent. In particular, show that $\widehat{\beta} \xrightarrow{P} a\beta$ and find a explicitly.

4. Let $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n) \sim P$. Assume that $X_i \in \{0, 1\}$ and let $\theta = \mathbb{E}[Y_1] - \mathbb{E}[Y_0]$ denote the causal effect. Assume there are no unmeasured confounders, that is, $X \amalg (Y_0, Y_1) \mid Z$. Let
$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(1, Z_i) - \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(0, Z_i)$$
where $\widehat{\mu}(x, z)$ is an estimate of $\mu(x, z)$. Suppose that
$$\sup_{x,z} |\widehat{\mu}(x, z) - \mu(x, z)| \xrightarrow{P} 0.$$
Show that $\widehat{\theta} \xrightarrow{P} \theta$.