STAT 2270 Fall 2020 HW1

Manuel Garcia Acosta

1.- a) $\mathbb{E}\left[\left[Y-\hat{f}(X)\right]^2\right] = \mathbb{E}\left[\left[f(X)-\hat{f}(X)+\varepsilon\right]^2\right]$

obs. ε independent of X

$$= \mathbb{E}\left[\left[f(X)-\hat{f}(X)\right]^2\right] + 2\,\mathbb{E}\left[\left[f(X)-\hat{f}(X)\right]\varepsilon\right]^{\nearrow 0} + \mathbb{E}\left[\varepsilon^2\right]$$

Obs. $\mathbb{E}[\varepsilon]=0$

$\Rightarrow Var(\varepsilon) = \mathbb{E}[\varepsilon^2]$

$$= \mathbb{E}\left[\left[\hat{f}(X)-f(X)\right]^2\right] + \mathbb{E}\left[\varepsilon^2\right] = \mathbb{E}\left[\left[\hat{f}(X)-f(X)\right]^2\right] + Var(\varepsilon)$$

Best error we can hope for is when first term is 0, since ε is the irreducible error

b) Notice the following

$$\mathbb{E}\left[\left[\hat{f}(X)-f(X)\right]^2\right] = \mathbb{E}\left[\left[\hat{f}(X)-\mathbb{E}[\hat{f}(X)]+\mathbb{E}[\hat{f}(X)]-f(X)\right]^2\right]$$

$$= \mathbb{E}\left\{\left[\hat{f}(X)-\mathbb{E}[\hat{f}(X)]\right]^2\right\} + \mathbb{E}\left[\left(\hat{f}(X)-\mathbb{E}[\hat{f}(X)]\right)^{\nearrow 0}\left(\mathbb{E}[\hat{f}(X)]-f(X)\right)\right]$$

$$+ \mathbb{E}\left[\left[\mathbb{E}[\hat{f}(X)]-f(X)\right]^2\right]$$

$$= Var\left(\hat{f}(X)\right) + Bias\left(\hat{f}(X)\right)^2$$

Finally we can rewrite

$$\mathbb{E}\left[\left[Y-\hat{f}(X)\right]^2\right] = Var\left(\hat{f}(X)\right) + Bias\left(\hat{f}(X)\right)^2 + Var(\varepsilon)$$

2.-

a) Let $p: \mathbb{R} \to \mathbb{R}$ be a non-zero function such that $p(X_i)=0$, $i \in \{1,\dots,n\}$. One such function could be

$$p(X) = \prod_{i=1}^{n} (X - X_i) \quad \text{where the } X_i\text{'s are those from } D_n$$

Notice $p(X_i) = 0$, $i \in \{1,\dots,n\}$, i.e. p is 0 over $\{X_1,\dots,X_n\}$. Now, consider a function $g: \mathbb{R} \to \mathbb{R}$ such that

$$g(X_i) = Y_i, \quad i \in \{1,\dots,n\}$$

Now, consider the regression function $\hat{f}_\lambda : \mathbb{R} \to \mathbb{R}$ defined as

$$\hat{f}_\lambda(X) = \lambda p(X) + g(X), \quad \lambda \in \mathbb{R} \text{ fixed } \lambda$$

By construction, it follows

$$\sum_{i=1}^{n} [Y_i - \hat{f}_\lambda(X_i)]^p = \sum_{i=1}^{n} [Y_i - (\lambda p(X_i) + g(X_i))]^p = \sum_{i=1}^{n} (Y_i - Y_i)^p = 0$$

The function $\hat{f}$ has perfect fit (zero loss). Moreover, this holds regardless of $\lambda \in \mathbb{R}$. If we consider the set of functions $F = \{\hat{f}_\lambda / \lambda \in \mathbb{R}\}$ we get an uncountable number of regression functions with perfect fit with respect to L_p loss for any p.

b) Consider a non-zero function $q: \mathbb{R} \to \mathbb{R}$ which has countable zeroes on X. This is $q(X_i) = 0$, $i \in \mathbb{N}$. Consider also a function $h: \mathbb{R} \to \mathbb{R}$ that satisfies $h(X_i) = Y_i$, $i \in \mathbb{N}$. Now consider the regression function $\hat{f}_\lambda : \mathbb{R} \to \mathbb{R}$

$$\hat{f}_\lambda(X) = \lambda q(X) + h(X), \quad \text{fixed } \lambda \in \mathbb{R}$$

Once again the regression function estimates will have perfect fit w.r.t. L_p loss

$$\sum_{i=1}^{\infty} [Y_i - \hat{f}_\lambda(X_i)]^p = \sum_{i=1}^{\infty} [Y_i - (\lambda q(X_i) + h(X_i))]^p = \sum_{i=1}^{\infty} [Y_i - Y_i]^p = 0$$

Considering $F = \{\hat{f}_\lambda / \lambda \in \mathbb{R}\}$ we get an uncountable number of regression functions with perfect fit with respect to L_p loss for any p.

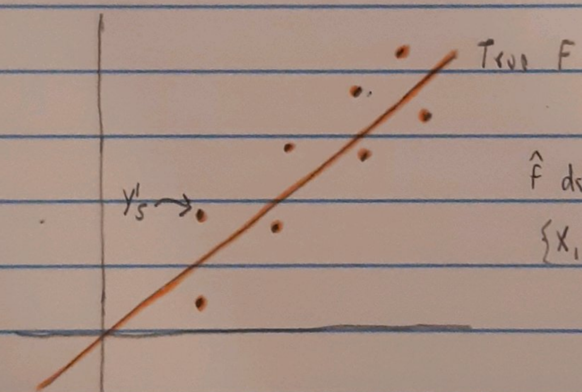
Exercise 2 Continued

c) Let $\varepsilon > 0$ and consider a function $\hat{f}: \mathbb{R} \to \mathbb{R}$ where

$$\hat{f}(x) = \begin{cases} f(x) & \text{if } x \notin \{X_1, \dots, X_n\} \\ Y_i & \text{if } x = X_i \end{cases}$$

Notice this regression estimate fits the data perfectly since $\hat{f}(X_i) = Y_i$, $i \in \{1, \dots, n\}$. Moreover, by construction $\hat{f}$ differs from f only on a finite set, i.e., $\hat{f}$ differs from f only over a set of measure zero. From all of the above $\hat{f}$ is an interpolator and the distance between f and $\hat{f}$ is less than ε using Lebesgue measure.

Regression function behave almost as the true f



$\hat{f}$ differs from f only on the finite set $\{X_1, \dots, X_n\}$. This set has measure 0.

3.- We use $Y = \beta^T X + \varepsilon$ and true model is $Y = \beta^T X + g(Z) + \varepsilon^*$. The form the error will take here is $\varepsilon = g(Z) + \varepsilon^*$. In order for our model to be appropriate we need to satisfy $\mathbb{E}[\varepsilon] = 0 \in \mathbb{R}^n$, $Var(\varepsilon) = I_n \sigma^2$. This will impose the conditions:

- $\mathbb{E}[g(Z) + \varepsilon^*] = 0_n$ the expected values should either cancel out or both expected values should be 0.
- $Var(g(Z) + \varepsilon^*) = I_n \sigma^2$

Obs 0_n is a vector in $\mathbb{R}^n$

- X needs to be independent of $g(Z)$ and ε^*
- If $\mathbb{E}[g(Z) + \varepsilon^*] \neq 0_n \in \mathbb{R}^n$ it won't be reasonable to assume unbiasedness
- If $g(Z)$ non random ε_i^* must be uncorrelated. If $g(Z)$ random $g(Z)_i + \varepsilon_i^*$ must be uncorrelated for all i.

- A strong requirement would be $g(Z) + \varepsilon^* \sim N(0, I\sigma^2)$ if $g(Z)$ random or $\varepsilon^* \sim N(0, I\sigma^2)$ if $g(Z)$ non-random