CS182: Introduction to Machine Learning: Homework #01

Due on Mar 7, 2023 at 23:59

Name: **Penghao Wang**
Student ID: 2021533138

Problem 1

1. As in least squares, we assume that the regression function $f(x)$ is approximately linear in such equation: $f(x) \approx x^T\beta$. As in EPE(f), we have that

$$
\begin{aligned}
\text{EPE}(f) &= \text{E}\left[L(Y, f(X))\right] \\
&= \text{E}(Y - f(X))^2 \\
&= \int [Y - f(X)]^2 \, Pr(dx, dy)
\end{aligned}
$$

With $f(X) \approx x^T\beta$, we will get that

$$
\beta = \left[E(XX^T)\right]^{-1} E(XY)
$$

.

2. Difference

   (a) As for nearest neighbor, it make the regression by taking the average of the nearest K neighbors as the predictions, which is same as take the max one in the K neighbors, and this is a non parametric method.

   (b) As for linear regression, it make the regression by estimates the regression function by minimizing the sum of the squared differences between the actual and predicted values. To make predictions, it just use the calculated parameters to make linear calculations to get the predictions. This is a parametric method.

   Above all, we can see that the main difference is that NN is a non parametric method, while linear regression is a parametric method. And when NN make predicts, it still need the training data to make predictions, but linear regression only use the calculated parameters to make predictions.

3. With given error loss $L(Y, f(X)) = |Y - f(X)|$, we have:

$$
\begin{aligned}
\text{EPE}(f) &= \text{E}\left[L(Y, f(X))\right] \\
&= \text{E}|Y - f(X)| \\
&= \int |Y - f(X)| Pr(dx, dy).
\end{aligned}
$$

We can conditioning on X, then we can write EPE as

$$
\text{EPE}(f) = \text{E}_X E_{Y|X}(|Y - f(X)| | X).
$$

Then we as it suffices to minimize EPE pointwise, we will get that:

$$
f(x) = argmin_c E_{Y|X}(|Y - c| | X = x).
$$

So the we get

$$
f(X) = \text{median}(Y|X = x).
$$

minimizes $EPE(f)$

Problem 2

To prove that $f(x) = \text{median}(Y|\ X{=}x)$ minimizes expected prediction error (EPE) using absolute error loss, we begin by defining the EPE as follows:

$EPE(f) = E[L(Y, f(X))] = E[|Y - f(X)|]$

From this equation, we know that minimizing EPE involves minimizing the expected value of the absolute difference between our predicted $(f(X))$ and actual (Y) values.

Let's assume that we have some fixed value of $X = x$ and denote the set of all observations such that $X = x$ as $\mathcal{N}(x)$. Using this notation, we can rewrite the above equation as:

$EPE(f|x) = E[|Y - f(x)|; |; X = x]$

Now, suppose that we select some other value $f^{'}(x)$. We can then write:

$EPE(f^{'}(x)|x) = E[|Y - f^{'}(x)|; |; X = x]$

Our goal is to show that $f(x) = \text{median}(Y|\ X{=}x)$ produces the smallest EPE. Let's start by splitting the absolute value into two separate cases: $Y < f(x)$ and $Y \geq f(x)$.

The expected prediction error only depends on how far the predictions are from the median value in each group:

$EPE(f|x) = E[|Y - f(x)|; |; X = x]$
$,,,,,,,,,, = E[(Y - f(x)) * I(Y < f(x)) + (f(x) - Y) * I(Y \geq f(x))]$

We can further simplify this expression by considering the expectation of each of these terms separately:

$E[Y - f(x)|X = x, Y < f(x)] + E[f(x) - Y|X = x, Y \geq f(x)]$

Both $E[Y - f(x)|X = x, Y < f(x)]$ and $E[f(x) - Y|X = x, Y \geq f(x)]$ are minimized if $f(x)$ is chosen such that an equal number of samples lie on either side of it. This, by definition, is the condition for choosing the median.

Hence, we have shown that $f(x) = \text{median}(Y|\ X{=}x)$ minimizes expected prediction error (EPE) using absolute error loss, which proves what was asked to be proven.