BZAN 615 - Homework 1

Jinyi Liu

Due: February 16, 2024

1 Absolute Error Loss

Show that the function $f_0(x)$ that minimizes expected prediction error with the absolute error loss $L(Y, f(X))) = |Y - f(X)|$ is given by

$$f_0(x) = \text{median}(Y \mid X = x)$$

where the median is defined by the equation $\mathbb{P}(Z \geq \text{median}(Z)) = 1/2$.

Proof. We have

$$\begin{aligned}
EPE(Y)(f) &= E|Y - f| \\
&= E(Y - f)_+ + E(f - Y)_+ \\
&= \int_0^\infty \mathbb{P}(Y - f \geq t)dt + \int_0^\infty \mathbb{P}(f - Y \geq t)dt \\
&= \int_0^\infty F_f(Y - t) + (1 - F_f(Y + t))dt
\end{aligned}$$

Then we have

$$\begin{aligned}
0 &= \partial EPE(Y)/\partial f \\
&= \int_0^\infty \mathbb{P}(f = Y - t) - \mathbb{P}(f = Y + t)dt,
\end{aligned}$$

which means that

$$\mathbb{P}(f \geq Y) = \mathbb{P}(f \leq Y) = 1 - \mathbb{P}(f \geq Y) \Rightarrow \mathbb{P}(Y \geq f) = \frac{1}{2},$$

i.e., the minimizer of $EPE(Y)(f)$ is the median of Y. $\square$

Proof. In the following we denote $F(f) \equiv F_Y(f) = \mathbb{P}(Y \leq f)$.

$$\begin{aligned}
EPE(Y)(f) &= E|Y - f| \\
&= E(Y-f)_+ + E(f-Y)_+ \\
&= \int_f^\infty (y-f)dF(y) + \int_{-\infty}^f (f-y)dF(y) \\
&= \int_f^\infty ydF(y) - f\int_f^\infty dF(y) + f\int_{-\infty}^f dF(y) - \int_{-\infty}^f ydF(y) \\
&= \int_f^\infty ydF(y) - f(1-F(f)) + F(f)f - \int_{-\infty}^f ydF(y)
\end{aligned}$$

Then we have

$$\begin{aligned}
\frac{\partial EPE(Y)}{\partial f} &= -fP(Y=f) - ((1-F(f) - fP(Y=f))) + P(Y=f)f + F(f) - fP(Y=f) \\
&= -fP(Y=f) - 1 + F(f) + P(Y=f)f + P(Y=f)f + F(f) - fP(Y=f) \\
&= -1 + F(f) + F(f) \\
&= 0 \Rightarrow F(f) = \frac{1}{2}.
\end{aligned}$$

□

2 Linear Function Space with Absolute Error Loss

Consider the case where input x is one-dimensional, and the sample size is 3. Let a be a real number. Let the data values be $(x_1, x_2, x_3) = (1, 1, 1)$ and $(y_1, y_2, y_3) = (1, 2, a)$

1. Find the linear function $f_\beta(x) = \beta x$ that minimizes the sample EPE under Squared error loss, call the minimizer β_{SE}.

   Proof. We have

$$\begin{aligned}
sEPE(f)(\beta) &= \hat{E}\left[(Y - f(X))^2\right] \\
&= \frac{1}{3}\left((1-\beta x)^2 + (2-\beta x)^2 + (a-\beta x)^2\right) \\
&= \frac{1}{3}\left(1 - 2\beta x + \beta^2x^2 + 4 - 4\beta x + \beta^2 x^2 + a^2 - 2a\beta x + \beta^2 x^2\right) \\
&= \frac{1}{3}\left(5 - 6\beta x + 3\beta^2x^2 + a^2 - 2a\beta x\right) \\
&= \frac{1}{3}\left(5 - 6\beta + 3\beta^2 + a^2 - 2a\beta\right).
\end{aligned}$$

Then we have

$$\frac{\partial sEPE(f)}{\partial \beta} = -6 + 6\beta - 2a = 0$$
$$\Rightarrow \beta_{SE} = \frac{3+a}{3}.$$

□

2. Find the linear function $f_\beta(x) = \beta x$ that minimizes the sample EPE under Absolute Error Loss, call the minimizer β_{AE}.

   Proof. We have

$$\begin{aligned} sEPE(f)(\beta) &= \hat{E}\left[|Y - f(X)|\right] \\ &= \frac{1}{3}\left(|1-\beta x| + |2-\beta x| + |a-\beta x|\right) \\ &= \frac{1}{3}\left(|1-\beta| + |2-\beta| + |a-\beta|\right). \end{aligned}$$

   By Problem 1, we have $\beta_{AE} = \text{median}(1, 2, a)$.

   □

3. Choose appropriate a to make $\beta_{SE} \neq \beta_{AE}$. Deduce that 'best' linear function depends on how we define 'best'.

   Proof. Choose $a = 2$, then we have $\beta_{SE} = \frac{5}{3}$ and $\beta_{AE} = 2$. Then we have $\beta_{SE} \neq \beta_{AE}$. □

3 Curse of Dimensionality

Let $\mathbf{x}_i$ be a random vector in p-dimensional space for each $i \in \{1, \cdots, n\}$. Each $\mathbf{x}_i$ is uniformly chosen from p dimensional unit cube (in other words components of the vector are indpendent and chosen from the interval $[0, 1]$.)

We define a neighborhood around zero

$$N(\delta) = \{\mathbf{x} \in \mathbb{R}^p : |\mathbf{x}_j| < \delta \text{ for all } j \in \{1, \cdots, p\}\}$$

In other words, $N(\delta)$ is a cube with side length δ whose one corner is at the origin.

1. What is the probability that $\mathbf{x}_1 \in N(\delta)$?

   Proof. We have

$$\begin{aligned} \mathbb{P}(\mathbf{x}_1 \in N(\delta)) &= \mathbb{P}(\mathbf{x}_{1,1} \in [0,\delta])\mathbb{P}(\mathbf{x}_{1,2} \in [0,\delta]) \cdots \mathbb{P}(\mathbf{x}_{1,p} \in [0,\delta]) \\ &= \delta^p. \end{aligned}$$

   □

2. What is the expected number of $\mathbf{x}_i$ that falls inside $N(\delta)$?

   Proof. We have

$$
\begin{aligned}
\mathbb{E}(\text{number of } \mathbf{x}_i \text{ in } N(\delta)) &= n\mathbb{P}(\mathbf{x}_1 \in N(\delta)) \\
&= n\delta^p.
\end{aligned}
$$

   □

3. Now choose $n = 1000, p = 100$ and $\delta = 0.01$. Compute the value in question 2

   Proof. We have

$$
\mathbb{E}(\text{number of } \mathbf{x}_i \text{ in } N(\delta)) = 1000 \times 0.01^{100} = 10^{-200}.
$$

   □

4. Comment on how answers in part 2 and 3 relate to the curse of dimensionality.

   Proof. The expected number of $\mathbf{x}_i$ that falls inside $N(\delta)$ is decreasing exponentially as the dimension p increases. This is the curse of dimensionality. □

4 Linear Algebra - Review

If $\mathbf{A}$ is a n by n matrix, and x is a n dimensional column vector. Define $\Psi(x) = x^T \mathbf{A} x$.

1. Write $\Psi(x)$ as a summation of x_i 's.

   Proof. We have

$$
\begin{aligned}
\Psi(x) &= x^T \mathbf{A} x \\
&= \sum_{i=1}^{n} x_i (\mathbf{A}x)_i \\
&= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} \mathbf{A}_{ij} x_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{A}_{ij} x_i x_j.
\end{aligned}
$$

   □

2. Compute $\frac{\partial \Psi}{\partial x_i}(x)$

Proof. We have

$$\begin{aligned}
\frac{\partial \Psi}{\partial x_i}(x) &= \frac{\partial}{\partial x_i}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}x_i x_j\right) \\
&= \sum_{j=1}^{n}\mathbf{A}_{ij}x_j + \sum_{j=1}^{n}\mathbf{A}_{ji}x_j \\
&= \sum_{j=1}^{n}\mathbf{A}_{ij}x_j + \sum_{j=1}^{n}\mathbf{A}_{ji}x_j \\
&= \sum_{j=1}^{n}(\mathbf{A}_{ij} + \mathbf{A}_{ji})x_j \\
&= \sum_{j=1}^{n}(\mathbf{A} + \mathbf{A}^T)_{ij}x_j.
\end{aligned}$$

□

3. Express $\nabla \Psi(x) = \frac{\partial}{\partial x}\Psi(x)$ in a vector notation.

Proof. We have

$$\nabla \Psi(x) = (\mathbf{A} + \mathbf{A}^T)x.$$

□