ST790 HW1

Steven Xu

September 5, 2019

1.

(a)

$$
\begin{aligned}
\min_a \mathbb{E}(X-a)^2 &= \min_a \mathbb{E}(X^2 - 2aX + a^2) \\
&= \mathbb{E}(X^2) + \min_a [a^2 - 2a\mathbb{E}(X)] \\
\frac{\partial a^2 - 2a\mathbb{E}(X)}{\partial a} &\overset{set}{=} 0 \\
\therefore\ [2a - 2\mathbb{E}(X)] &= 0 \\
a &= \mathbb{E}(X) \\
\because\ \frac{\partial^2 [a^2 - 2a\mathbb{E}(X)]}{\partial a^2} &= 2 > 0 \\
\therefore\ \min_a \mathbb{E}(X-a)^2 &= \mathbb{E}[X - \mathbb{E}(X)]^2
\end{aligned}
$$

(b)

$$
\begin{aligned}
\min_a \mathbb{E}|X-a| &= \min_a \left\{ \int_{-\infty}^{a} (a-x) f(x) dx + \int_{a}^{\infty} (x-a) f(x) dx \right\} \\
\frac{\partial \mathbb{E}|X-a|}{\partial a} &= \frac{\partial}{\partial a} \int_{-\infty}^{a} (a-x) f(x) dx + \frac{\partial}{\partial a} \int_{a}^{\infty} (x-a) f(x) dx \\
&= (a-a)f(a) + \int_{-\infty}^{a} f(x) dx + (a-a) f(a) - \int_{a}^{\infty} f(x) dx \text{ (Leibniz Rule)} \\
&\overset{set}{=} 0 \\
\therefore\ F(a) - [1 - F(a)] &= 0 \\
F(a) &= \frac{1}{2} \\
\therefore\ a &= m \\
\because\ \frac{\partial^2 \mathbb{E}|X-a|}{\partial a^2} &= 2f(a) \geq 0 \\
\therefore\ \min_a \mathbb{E}|X-a| &= \mathbb{E}|X-m|
\end{aligned}
$$

2.

(a)

$$
\begin{aligned}
\because\ X_1, X_2, \ldots, X_n &\overset{iid}{\sim} \text{Ber}(p) \\
\therefore\ \mathcal{L}(p; \mathbf{x}) &= \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}
\end{aligned}
$$

$$
\begin{aligned}
&= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \\
p(p|\mathbf{x}) &\propto \mathcal{L}(p;\mathbf{x})\pi(p) \\
&= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1} \\
&\propto p^{\alpha+\sum_{i=1}^n x_i-1}(1-p)^{\beta+n-\sum_{i=1}^n x_i-1}
\end{aligned}
$$

We recognize that this is the kernel of $Beta(\alpha+\sum_{i=1}^n x_i, \beta+n-\sum_{i=1}^n x_i)$

(b)

By the property of Beta distribution, $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ if $X \sim Beta(\alpha,\beta)$.

$$
\begin{aligned}
\therefore\ \hat{p}_{\text{Bayes}} &= \mathbb{E}(p|\mathbf{x}) \\
&= \frac{\alpha+\sum_{i=1}^n x_i}{\alpha+\beta+n}
\end{aligned}
$$

(c)

Let $Y = \sum_{i=1}^n X_i$, then $Y \sim Bin(n,p)$. Under the squared error loss

$$
\begin{aligned}
R(p,\hat{p}_{Bayes}) &= \mathbb{E}(p-\hat{p}_{Bayes})^2 \\
&= \mathbb{E}(p-\frac{\sum_{i=1}^n x_i+\alpha}{\alpha+\beta+n})^2 \\
&= \mathbb{E}\left[p^2-\frac{2p(y+\alpha)}{\alpha+\beta+n}+\frac{(y+\alpha)^2}{(\alpha+\beta+n)^2}\right] \\
&= p^2-\frac{2np^2+2p\alpha}{\alpha+\beta+n}+\frac{\mathbb{E}(y^2)+2\alpha\mathbb{E}(y)+\alpha^2}{(\alpha+\beta+n)^2} \\
&= p^2-2\frac{p(np+\alpha)}{\alpha+\beta+n}+\frac{np(1-p)+(np)^2+2\alpha np+\alpha^2}{(\alpha+\beta+n)^2} \\
&= \left(p-\frac{np+\alpha}{\alpha+\beta+n}\right)^2+\frac{np(1-p)}{(\alpha+\beta+n)^2}
\end{aligned}
$$

(d)

When $\alpha=\beta=\sqrt{n/4}$

$$
\begin{aligned}
\hat{p}_{\text{Bayes}} &= \frac{\sqrt{n/4}+\sum_{i=1}^n x_i}{\sqrt{n}+n} \\
R(p,\hat{p}_{\text{Bayes}}) &= \left(p-\frac{np+\sqrt{n/4}}{\sqrt{n}+n}\right)^2+\frac{np(1-p)}{(\sqrt{n}+n)^2} \\
&= \left(\frac{\sqrt{n}p+np-np-\sqrt{n/4}}{\sqrt{n}+n}\right)^2+\frac{np(1-p)}{(\sqrt{n}+n)^2} \\
&= \left[\frac{\sqrt{n}(p-1/2)}{\sqrt{n}+n}\right]^2+\frac{np(1-p)}{(\sqrt{n}+n)^2} \\
&= \frac{n(p-1/2)^2}{(\sqrt{n}+n)^2}+\frac{np(1-p)}{(\sqrt{n}+n)^2} \\
&= \frac{n/4}{(\sqrt{n}+n)^2} \\
&= \frac{n}{4(\sqrt{n}+n)^2}
\end{aligned}
$$

3.

(a)

Linear regression:

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\
&= \begin{pmatrix} 1 & x_i \end{pmatrix} \boldsymbol{\beta} + \epsilon_i \\
\therefore\ \hat{f}(x_0) &= \begin{pmatrix} 1 & x_0 \end{pmatrix} \hat{\boldsymbol{\beta}} \\
&= \begin{pmatrix} 1 & x_0 \end{pmatrix} (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \begin{pmatrix} 1 & x_0 \end{pmatrix} \frac{1}{N\sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \begin{pmatrix} \sum_{i=1}^N x_i^2 & -\sum_{i=1}^N x_i \\ -\sum_{i=1}^N x_i & N \end{pmatrix} \mathbf{X}^T\mathbf{y} \\
&= \frac{1}{N\sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \begin{pmatrix} \sum_{i=1}^N x_i^2 - x_0 \sum_{i=1}^N x_i & Nx_0 - \sum_{i=1}^N x_i \end{pmatrix} \mathbf{X}^T\mathbf{y} \\
&= \sum_{i=1}^N \frac{1}{N\sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \left[\sum_{i=1}^N (x_i^2 - x_0 x_i) + N x_0 x_i - x_i \sum_{i=1}^N x_i \right] y_i \\
\therefore\ \ell_i(x_0, \mathcal{X}) &= \frac{1}{N\sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \left[\sum_{i=1}^N (x_i^2 - x_0 x_i) + N x_0 x_i - x_i \sum_{i=1}^N x_i \right]
\end{aligned}$$

K-nearest neighbors:

$$\begin{aligned}
\hat{f}(x_0) &= k^{-1} \sum_{x_i \in N(x_0)}^N y_i \\
\therefore\ \ell_i(x_0, \mathcal{X}) &= \mathbf{1}(x_i \in N(x_0))
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbb{E}_{\mathcal{Y}|\mathcal{X}}(f(x_0) - \hat{f}(x_0))^2 &= \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left(f(x_0) - \sum_{i=1}^N \ell_i(x_0, \mathcal{X}) y_i \right)^2 \\
&= f(x_0)^2 - 2f(x_0) \sum_{i=1}^N \ell_i(x_0, \mathcal{X}) \mathbb{E}(y_i|\mathcal{X}) + \mathbb{E}\left[\left(\sum_{i=1}^N \ell_i(x_i, \mathcal{X}) y_i \right)^2 |\mathcal{X} \right] \\
&= f(x_0)^2 - 2f(x_0) \sum_{i=1}^N \ell_i(x_0, \mathcal{X}) f(x_i) + \mathrm{Var}\left[\sum_{i=1}^N \ell_i(x_0, \mathcal{X}) y_i | \mathcal{X} \right] + \mathbb{E}\left(\sum_{i=1}^N \ell_i(x_i, \mathcal{X}) y_i | \mathcal{X} \right)^2 \\
&= f(x_0)^2 - 2f(x_0) \sum_{i=1}^N \ell_i(x_0, \mathcal{X}) f(x_i) + \sigma^2 \sum_{i=1}^N \ell_i(x_0, \mathcal{X})^2 + \left[\sum_{i=1}^N \ell_i(x_0, \mathcal{X}) f(x_i) \right]^2 \\
&= \left[f(x_0) - \sum_{i=1}^N \ell_i(x_0, \mathcal{X}) f(x_i) \right]^2 + \sigma^2 \sum_{i=1}^N \ell_i(x_0, \mathcal{X})^2
\end{aligned}$$

(c)

By the law of iterated expectation,

$$\begin{aligned}
\mathbb{E}_{\mathcal{Y},\mathcal{X}}(f(x_0) - \hat{f}(x_0))^2 &= \mathbb{E}[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}(f(x_0) - \hat{f}(x_0)^2)] \\
&= f(x_0)^2 - 2f(x_0) \sum_{i=1}^N \mathbb{E}[\ell_i(x_0, \mathcal{X}) f(x_i)] + \sigma^2 \sum_{i=1}^N \mathbb{E}\left[\ell_i(x_0, \mathcal{X})^2 \right] + \mathbb{E}\left\{ \left[\sum_{i=1}^N \ell_i(x_0, \mathcal{X}) f(x_i) \right]^2 \right\}
\end{aligned}$$

$$= \left\{ f(x_0) - \sum_{i=1}^{N} \mathbb{E}[\ell_i(x_0, \mathcal{X}) f(x_i)] \right\}^2 + \sigma^2 \sum_{i=1}^{N} \mathbb{E}\left[\ell_i(x_0, \mathcal{X})^2\right] + \mathrm{Var}\left[\sum_{i=1}^{N} \ell_i(x_0, \mathcal{X}) f(x_i)\right]$$

(d)

$$\begin{aligned}
\mathrm{Bias}^2_{\mathcal{Y},\mathcal{X}}(\hat{f}(x_0)) &= \mathbb{E}[\mathrm{Bias}^2_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))] - \mathrm{Var}[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))] \\
\mathrm{Var}_{\mathcal{Y},\mathcal{X}}(\hat{f}(x_0)) &= \mathbb{E}[\mathrm{Var}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))] + \mathrm{Var}[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))]
\end{aligned}$$