STAT 5701: Statistical Computing

Homework 1

Jingxiang Li

September 25, 2015

Problem 1.a

The optimal $f(x)$ is $f(x) = E(y|x)$

Proof. The expected loss is

$$\begin{aligned} E_{(x,y)}[l(f(x), y)] &= \int_x \int_y l(f(x), y) p(y|x) dy p(x) dx \\ &= \int_x \int_y (f(x) - y)^2 p(y|x) dy p(x) dx \end{aligned}$$

Here we minimize the expected loss

$$\min_f \mathrm{E}_{(x,y)}[l(f(x), y)] \Leftrightarrow \min_f \int_y (f(x) - y)^2 p(y|x) dy$$

To minimize the expected loss, we take the derivative with respect to f, and set it to 0.

$$\begin{aligned} &\frac{\partial}{\partial f} \int_y (f(x) - y)^2 p(y|x) dy = 0 \\ \Rightarrow &\int_y \frac{\partial}{\partial f} (f(x) - y)^2 p(y|x) dy = 0 \\ \Rightarrow &\int_y (f(x) - y) p(y|x) dy = 0 \\ \Rightarrow &\int_y f(x) p(y|x) dy = \int_y y p(y|x) dy \\ \Rightarrow &f(x) = E(y|x) \end{aligned}$$

Q.E.D

Problem 2.a

Proof. The expected loss for f is

$$L(f) = p(f(x) \neq y) = \mathrm{E}_{(x,y)}[I(f(x) \neq y)] = \int_x \sum_y I(f(x) \neq y) p(y|x) p(x) dx$$

Here we minimize the expected loss with respect to f

$$\begin{aligned} &\min_f \int_x \sum_y I(f(x) \neq y) p(y|x) p(x) dx \\ \Leftrightarrow &\min_f \sum_y I(f(x) \neq y) p(y|x) \\ \Leftrightarrow &\min_f I(f(x) \neq 1) p(1|x) + I(f(x) \neq -1) p(0|x) \end{aligned}$$

Notice that $p(0|x) > 0$, $p(1|x) > 0$ and $p(0|x) + p(1|x) = 1$.
Considering $I(f(x) \neq 1)$ and $I(f(x) \neq 0)$, there will be exactly one has value 1, and another one 0.
Hence once $p(1|x) > p(0|x)$, the optimal f^* must satisfy $f^*(x) = 1$; otherwise $f^*(x) = -1$
i.e.

$$f^*(x) = \begin{cases} +1, & \text{if } p(1|x) > 1/2 \\ -1, & \text{otherwise} \end{cases}$$

Since f^* is optimal, $L(f^*) \leq L(f)$
Q.E.D.