Notes for *Elements of Statistical Learning*

Sebastian Claici

November 25, 2024

1 Chapter 2

The goal is to minimize the expected prediction error:

$$\begin{aligned}\mathrm{EPE}(f) &= \mathrm{E}\left(Y - f(X)\right)^2 \\ &= \int [y - f(x)]^2 p(x, y) \,\mathrm{d}x \,\mathrm{d}y\end{aligned}$$

If we break down the expectation as $E_{X,Y} = E_X E_{Y|X=x}$ we can rewrite this as

$$\begin{aligned}\mathrm{EPE}(f) &= \mathrm{E}_X \mathrm{E}_{Y|X=x}(Y - f(X))^2 \\ &= \int_X \int_Y [y - f(x)]^2 p(y|x) p(x) \,\mathrm{d}y \,\mathrm{d}x \\ &= \int_X p(x) \left(\int_Y [y - f(x)]^2 p(y|X = x) \,\mathrm{d}y \right) \mathrm{d}x\end{aligned}$$

We have moved the dependence on $p(x)$ outside the inner expectation. Since f is unconstrained, we can solve for the optimal f pointwise. That is:

$$\arg\min_f \mathrm{EPE}(f) = \arg\min_c \int_Y [y - c]^2 p(y|X = x) \,\mathrm{d}y$$

Differentiating wrt c and using the fact that

$$\int_Y y p(y|X = x) \,\mathrm{d}y = \mathrm{E}(Y|X = x)$$

gives us (2.13) in the book.