Week14-notes

Monday, 19 April 2021 12:15

Extra Exercises

So $EPE(C) \leq EPE(g)$ for all functions g.

2)

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

$$= E[(Y - f(X))^2]$$

$$= \int_X \int_Y (y - f(x))^2 \, p(x,y) \, dy dx$$

$$= \int_X \int_Y (y - f(x))^2 \, \overbrace{p(x) p(y|x)}^{} \, dy dx$$

$$= \int_X \underbrace{\left(\int_y (y - f(x))^2 \, p(y|x) dy \right)}_{} p(x) dx$$

$$\underbrace{= \int_X E_{\boxed{Y|X=x}}[(Y - f(X))^2 | X=x] \; p(x) dx}_{}$$

$$\int_X 1 \cdot p(x) dx = \int_X p(x) dx = 1$$

$$= \boxed{E_X}[E_{Y|X}[(Y - f(X))^2 | X=x]] = E_{Y|X}[(Y - f(X))^2 | X=x] \cdot \underbrace{\overset{\shortparallel}{E_X[1]}}_{1}$$

We are looking for a function f that min. $EPE(f)$ given the data (i.e, $X=x$), then $EPE(f)$ becomes

$$\boxed{E_{Y|X=x}[(Y - f(X))^2 | X=x]}$$

We want a func. f that minimizes this expression, that is (by def)

$$\Rightarrow \boxed{f(x) = \arg\min_c \underline{E_{Y|X=x}[(Y-c)^2 | X=x]}}$$

b) We want the value of c that minimizes

$$L = E_{Y|X=x}[(Y-c)^2 | X=x]$$

$$= E_{Y|X=x}[Y^2 - 2cY + c^2 | X=x]$$

$$= E_{Y|X=x}[Y^2 | X=x] - 2c E_{Y|X=x}[Y | X=x] + c^2$$

Take the first derivative wrt c.

$$\frac{\partial L}{\partial c} = 0 - 2E_{Y|X=x}[Y | X=x] + 2c$$

Set this equal to 0 and solve for c.

$$\underline{-2E_{Y|X=x}[Y | X=x]} + 2C = 0$$

$$C = E_{Y|X=x}[Y | X=x]$$

Check that it is a minimum. 2nd derivative test.

$$\frac{\partial^2 L}{\partial c^2} = 2 > 0$$

So $\underline{c = E_{Y|X=x}[Y | X=x]}$ is the minimizer of L.

c)

Consider a model

$$Y = g(X) + \mathcal{E},$$

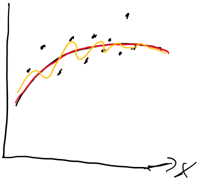
where $E[\mathcal{E}] = 0$.

We know from b) that

$c = E_{Y|X=x}[Y|X=x]$ is the minimizer of $EPE(f)$, so we plug in the given expression of Y into this solution.

$$c = E_{Y|X=x}[Y|X=x]$$
$$= E_{Y|X=x}[g(X)+\varepsilon \mid X=x]$$
$$= E_{Y|X=x}[g(X)|X=x] + E_{Y|X=x}[\varepsilon|X=x]$$
$$= E_{Y|X=x}[\boxed{g(x)}] + E_{Y|X=x}[\varepsilon]$$
$$= g(x)\, E_{Y|X=x}[1] + 0$$
$$= g(x) \cdot 1$$

So, $f(\cdot)$ is the optimal predictor when $f(\cdot) = g(\cdot)$



d) $V(\varepsilon) = \sigma^2$, ε ind. of X.

law of total Expectation.
$E[Y] = E[E[Y|X]]$

$$EPE(f) = E[(Y - f(X))^2]$$
$$= E[((Y - E[Y]) + (E[Y] - f(X)))^2]$$
$$= E[(Y - E[Y])^2 + (E[Y] - f(X))^2 + 2(Y - E[Y])(E[Y] - f(X))]$$
$$= \underbrace{E[(Y - E[Y])^2]} + E[(E[Y] - f(X))^2] + 2E[(Y - E[Y])(E[Y] - f(X))]$$
$$= Var(Y) + E[(E[Y] - f(X))^2] + 2E_X[E[(Y - E[Y])(E[Y] - f(X))|X]]$$
$$= \text{———} \; || \; \text{———} + 2E_X[(E[Y] - f(X))\underbrace{E[(Y - E[Y])|X]}_{0}]$$
$$= Var(Y) + E[(E[Y] - f(X))^2]$$
$$= Var(f(X) + \varepsilon) + E[(E(Y) - f(X))^2]$$
$$= Var(f(X)) + Var(\varepsilon) + E[(E(Y) - f(X))^2] \quad 0$$
$$= Var(f(X)) + \sigma^2 + E[(E[Y] - f(X))^2]$$

$$-E[E[Y]] + E[Y|X]$$
$$= -E[Y] + E[Y|X]$$
$$= -E[E[Y|X]] + E[Y|X]$$
$$= E[\underbrace{-E[Y|X] + E[Y|X]}_{0}]$$
$$= 0$$

Could have been showed easier.

The last term will be zero when $E[Y] = f(X)$, so, the lower bound is $Var(f(X)) + \sigma^2$

Extra 3

a)

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

$$= E[1 - I_{\{f(x)\}}(y)]$$

here

$$\underset{\uparrow}{I_{\{f(x)\}}}(y) = \begin{cases} 0 & \text{if } y \neq f(x) \\ 1 & \text{if } y = f(x) \end{cases} = I\{y = f(x)\}$$

Indicator function

$$EPE(f) = \int_X \int_Y (1 - I_{\{f(x)\}}(y))\, p(x, y)\, dy\, dx$$

$$= \int_X \left[\int_Y (1 - I_{\{f(x)\}}(y))\, p(y|x) \overset{dy}{\downarrow} \right] p(x)\, dx$$

Look at the inner part

$$E[I_{\{f(x)\}}(y)] = Pr(Y = f(X) \mid X = x)$$

$$\int_Y (1 - I_{\{f(x)\}}(y))\, p(y|x)\, dy$$

$$= \int_Y p(y|x)\, dy - \int_Y I_{\{f(X)\}}(y)\, p(y|x)\, dy$$

$$= \underline{1} - \underline{Pr(Y = f(X) \mid X = x)}$$

$$\rightarrow \int_X (1 - Pr(Y = f(x) \mid X = x))\, p(x)\, dx$$

b)

We have that $EPE(f) = \int_X \boxed{\overbrace{(1 - Pr(Y = \boxed{f(x)} \mid X = x)}^{?}}\, \boxed{p(x)}\, dx$

We are looking for a function f that minimizes this expression, which is by definition,

$$f(x) = \arg\min_{k} \{1 - Pr(Y = k \mid X = x)\}$$

$$= \arg\max_{k} \Pr(Y=k \mid X=x)$$

where $k \in \{0,1\}$

since binary classification, we have that $f(X)$ is either 0 or 1.

we are maxizing $\Pr(Y=k \mid X=x)$, so if $\Pr(Y=0 \mid X=x) < \Pr(Y=1 \mid X=x)$, then $k=1$. similarily, if

$$\Pr(Y=1 \mid X=x) < \Pr(Y=0 \mid X=x),$$

then $k=0$.

Furthermore, we have that

$$\Pr(Y=0 \mid X=x) + \Pr(Y=1 \mid X=x) = 1$$

So, the decision boundary is at

$$\Pr(Y=0 \mid X=x) = P(Y=1 \mid X=x) = 0.5$$

Therefore,

$$f(x) = \begin{cases} 1 & \text{if } \Pr(Y=1 \mid X=x) > 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

C) Extend the results to K classification.

Intuitively

$$f(x) = \begin{cases} K-1 & \text{if } K-1 = \underset{c}{\arg\max}\ \Pr(Y=c \mid X=x) \\ K-2 & \text{if } K-2 = \underset{c}{\arg\max}\ \Pr(Y=c \mid X=x) \\ \vdots & \vdots \\ 1 & 1 = \text{—}\ \text{''}\ \text{—} \\ 0 & \text{otherwise.} \end{cases}$$

d) Assume that $k_{opt} = \underset{k}{\arg\max}\ \Pr(Y=k \mid X=x)$. Then, we get an error when $Y \neq k_{opt}$. the probability of this happening is $1 - \Pr(Y=k_{opt} \mid X=x)$, but this is the same as

$$\Downarrow$$

$$1 - \max_{k} \Pr(Y=k \mid X=x).$$