SET-UP

$\mathbf{X} \in \mathbb{R}^p$ real valued RV of covariates

$Y \in \mathbb{R}$ response

$P(\mathbf{X}, Y)$

<u>Aim</u>: find $f(\mathbf{X})$ to predict $Y$

$L(Y, f(\mathbf{X}))$ is our <u>loss function</u>

Find $f(\mathbf{X})$ to minimize the $EPE(f)$
↑ expected prediction error

$EPE(f) = E_{X,Y}\left(L(Y, f(\mathbf{X}))\right)$ assume $Y$ and $\mathbf{X}$ continuous

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L(y, f(x))\, p(x,y)\, dx\, dy$$

↑ usually $f(x,y)$ but we use $f$ for our func. of interest

Q: why not $\iint L(y, f(x))\, p(f(x), y)\, d f(x)\, dy$ ?

"Law of the unconscious statistician" $E(g(\mathbf{X}, Y)) = \iint g(x,y)\, p(x,y)\, dx\, dy$

With squared loss: $L(Y, f(X)) = (Y - f(X))^2$ $L_2$-loss

realy of integrals

$$EPE(f) = E\left[Y - f(X))^2\right] = \int_x \int_y (y - f(x))^2 \, p(x,y) \, dy \, dx$$

constant wrt y

$$= \iint_{x\,y} (y - f(x))^2 \, p(y|x) \, p(x) \, dy \, dx = \int_x p(x) \cdot \underbrace{\int_y (y - f(x))^2 p(y|x) \, dy}_{} \, dx$$

$$E_X\left[ E_{Y|X}\left[(Y - f(X))^2 \mid X\right]\right]$$

What does this mean? May optimize $f(X)$ at each $X = x$ at aline (pointwise)

$EPE(f)$ at $X = x$ found by minimizing with respect to $f(x)$

$$E_{Y|X}\left( (Y - f(X))^2 \mid X \right) = \int_y (y - f(x))^2 \, p(y|x) \, dy$$

$$0 = \frac{d}{df} \int_y (y - f(x))^2 \, p(y|x) \, dy \overset{*}{=} \int \frac{d}{df} (y - f)^2 \, p(y|x) \, dy$$

$$= \int -2(y - f) \, p(y|x) \, dy = -2E(Y \mid X = x) + 2 f(x) \overbrace{\int_y p(y|x) \, dy}^{1}$$

$$E(Y \mid X = x) = f(x)$$

$*$ if we may change the order of $\int$ and $\frac{d}{df}$

But, do we know $E(Y|X=x)$?

Sometimes? What if $\begin{bmatrix} Y \\ X \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix}, \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \right)$?

Find $E(Y|\mathbf{X}=x)$

From TMA4267 we know that $Y|\mathbf{X}=x \sim N$ with

$$E(Y|X=x) = \mu_Y + \Sigma_{YX}\Sigma^{-1}(x-\mu_X)$$

and $Var(Y|X=x) = \Sigma_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$

What can we observe? $E(Y|\mathbf{X}=x)$ is a <u>linear</u> funct in $x$ and $Var(Y|X=x)$ is not dept of $x$

$\Rightarrow$ "main" reason for why we love MLR so much!

Rule

$$\frac{d\, X^T\beta}{d\beta} = X$$

$$d\ \frac{\beta^T X^T X \beta}{d\beta} = 2X^T X\beta$$