Basic Plot

pair of random variables X, Y

Q: given X, what do we know about Y?

conditional density $f(y|x)$

Describe expected value of Y, conditioned on $X=x$ as a function of x.

Part 1 $\mu_{Y|x} = E(Y|X=x) = \int y f(y|x)\, dy$

an expression of $\mu_{Y|x}$ is a "regression equation"

Part 2 Linear regression

Usual assumption is $\mu_{Y|x}$ is a linear function of x.

i.e. $\boxed{\mu_{Y|x} = \alpha + \beta x}$

Q: how do we determine α & β?

can express α & β in terms of $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \sigma_{X,Y}$

$\sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]$

<u>Digression:</u> express α, β in terms of these

trick:

mult by $g(x)$ & $\int dx$

$$\mu_{y|x} = \alpha + \beta x$$

mult by $x g(x)$ $\int dx$

let $g(x)$ = marginal density for x

$$g(x) = \int f(x,y)\,dy$$

$$\mu_{y|x}\, g(x) = \alpha g(x) + \beta x g(x)$$

$$\mu_{y|x} = \int y f(y|x)\,dy$$

$$\iint y \underbrace{f(y|x)\,g(x)}_{f(x,y)}\,dx\,dy = \alpha \underbrace{\int g(x)\,dx}_{1} + \beta \underbrace{\int x g(x)\,dx}$$

$$= E[y] = \mu_y$$

$$\int x g(x)\,dx = \int x\left(\int f(x,y)\,dy\right)dx$$

$$= \iint x f(x,y)\,dx\,dy$$

$$= E[x] = \mu_x$$

$$\boxed{\mu_y = \alpha + \beta \mu_x}$$

$$\mu_{y|x}\, x\, g(x) = \alpha\, x\, g(x) + \beta x^2 g(x)$$

$$\iint x y \underbrace{f(y|x) g(x)}_{f(x,y)}\, dx\, dy = \alpha \int x g(x)\, dx + \beta \int x^2 g(x)\, dx$$

$$= \alpha \iint x\, f(x,y)\, dx\, dy + \beta \iint x^2 f(x,y)\, dx\, dy$$

$$\underbrace{E[XY]}_{\sigma_{x,y} + \mu_x \mu_y} = \alpha \underbrace{E[X]}_{= \mu_x} + \beta \underbrace{E[X^2]}_{\sigma_x^2 + \mu_x^2}$$

$$\sigma_{x,y} + \mu_x \mu_y = \alpha \mu_x + \beta \sigma_x^2 + \beta \mu_x^2$$

$$\uparrow$$

$$\mu_y = \alpha + \beta \mu_x$$

$$\sigma_{x,y} + \cancel{\alpha \mu_x} + \cancel{\beta \mu_x^2} = \cancel{\alpha \mu_x} + \beta \sigma_x^2 + \cancel{\beta \mu_x^2}$$

$$\sigma_{x,y} = \beta \sigma_x^2 \qquad \beta = \frac{\sigma_{x,y}}{\sigma_x^2}$$

$$\alpha = \mu_y - \frac{\sigma_{x,y}}{\sigma_x^2} \mu_x$$

$$\mu_{y|x} = \alpha + \beta x = \mu_y - \frac{\sigma_{x,y}}{\sigma_x^2} \mu_x + \frac{\sigma_{x,y}}{\sigma_x^2} x$$

$$\boxed{\mu_{y|x} = \mu_y + \frac{\sigma_{x,y}}{\sigma_x^2}(x - \mu_x)}$$

we just showed if $\mu_{y|x}$ is a lnr function of x, then it is this lnr function.

Natural guess: given samples (x_i, y_i) then via "method of moments"

we can use point estimates

$S_{x,x}$, $\bar{X}$, $\bar{Y}$, $S_{x,y}$ for

σ_x^2, μ_x, μ_y, $\sigma_{x,y}$ to get estimates

for α, β

$$\text{e.g. } \beta = \frac{\sigma_{x,y}}{\sigma_x^2} \approx \frac{S_{x,y}}{S_{x,x}}$$

Part 3 "Least squares"

<u>Problem</u>: given pairs (x_i, y_i) find an eqn for a line $y = \alpha + \beta x$ s.t.

$$\sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2 \text{ minimized}$$

do same calculus... get.

$$\beta = \frac{S_{x,y}}{S_{x,x}}$$ where $S_{x,y} = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$

$S_{x,x} = \frac{1}{n}\sum(x_i - \bar{x})^2$

$$y \approx \bar{y} + \frac{S_{x,y}}{S_{x,x}}(x - \bar{x})$$