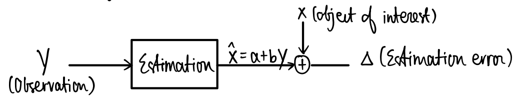
Linear Least Square Error (LLSE) Estimation



Goal / Find the LLSE estimate of X given Y.

Assume joint statistic of X, Y are known, i.e. $f(x,y)$ known.

$L[X|Y] = \hat{X} = a + bY$, where a, b are chose to min $\mathbb{E}[\Delta^2]$

$\underbrace{\mathbb{E}[\Delta^2]}_{\xi(a,b)} = \mathbb{E}[(X-\hat{X})^2] = \mathbb{E}[(X-a-bY)]$

$\frac{\partial \xi(a,b)}{\partial a} = 0 \Rightarrow \frac{\partial}{\partial a}\mathbb{E}[\cdot] = \mathbb{E}[\frac{\partial}{\partial a}(\cdot)] = \mathbb{E}[-2(X-a-bY)] = 0$

$\mathbb{E}[X] = a + b\,\mathbb{E}[Y]$

$\mathbb{E}[X] = \mathbb{E}[\hat{X}]$

$\boxed{\mathbb{E}[\Delta] = 0}$ Error is <u>unbiased</u>

$\frac{\partial \xi(a,b)}{\partial b} = 0 \Rightarrow -2\,\mathbb{E}[(X-a-bY)Y] = 0$

$\mathbb{E}[XY - (a-bY)Y] = 0$

$\mathbb{E}[XY - \hat{X}Y] = 0$

$\mathbb{E}[(X-\hat{X})Y] = 0$

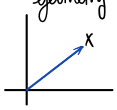
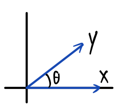
$\boxed{\mathbb{E}[\Delta Y] = 0}$ Error is <u>uncorrelated</u> w/ the observation

$$\boxed{L[X|Y] = \mathbb{E}[X] + \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathbb{E}[Y])}$$

Geometry of Random Vectors

↳ can picture RVs as vectors

Assume that X, Y are zero mean RVs w/ finite second moments, i.e. $\mathbb{E}[X^2] < \infty$, $\mathbb{E}[Y^2] < \infty$, then we have the following association:

Prob concept	Geometry
① RV X	
② RVs X & Y	
③ $\mathbb{E}[XY]$	$\langle \vec{x}, \vec{y} \rangle = \|\vec{x}\| \cdot \|\vec{y}\| \cos\theta$
3a) $\mathbb{E}[XY] = 0$	$\theta = \frac{\pi}{2}$ (orthogonal)
④ $\mathbb{E}[X^2]$	$\langle \vec{x}, \vec{x} \rangle = \|\vec{x}\|^2$ norm
⑤ $\rho = \frac{\mathbb{E}[XY]}{\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}}$ correlation	$\frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\|\|\vec{y}\|} = \cos\theta$ When $\theta = \frac{\pi}{2}$, X, Y are uncorrelated ($X \perp\!\!\!\perp Y$)

Geometric Derivation of LLSE

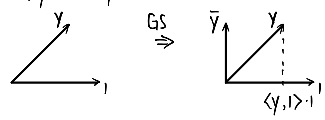
$L[X|Y] = a + bY$

span$\{1, Y\}$

Basis $\{1, Y\}$ (not an orthogonal basis)

GS $\Rightarrow \{1, \bar{Y}\}$ (orthonormal basis)

$\langle 1, Y\rangle = \mathbb{E}[Y] \neq 0$



$$\bar{Y} = \frac{Y - \langle Y, 1\rangle \cdot 1}{|Y - \langle Y, 1\rangle \cdot 1|}$$

$$= \frac{Y - \mathbb{E}[Y]}{\underbrace{|Y - \mathbb{E}[Y]|}_{\sqrt{\mathrm{Var}(Y)}}}$$

$$\bar{Y} = \frac{Y - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}(Y)}}$$

$$L[X|Y] = P_{\{1,Y\}} X$$

$$= P_{\{1,\bar{Y}\}} X$$

$$= P_1 X + P_{\bar{Y}} X$$

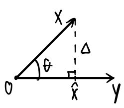
$$= \langle X, 1\rangle \cdot 1 + \langle X, \bar{Y}\rangle \bar{Y}$$

$$= \mathbb{E}[X] + \underbrace{\mathbb{E}[X\bar{Y}]}\cdot \bar{Y}$$

$$= \mathbb{E}[X] + \frac{\mathbb{E}[X(Y - \mathbb{E}[Y])]}{\sqrt{\mathrm{Var}(Y)}} \frac{(Y - \mathbb{E}[Y])}{\sqrt{\mathrm{Var}(Y)}}$$

$$= \mathbb{E}[X] + \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathbb{E}[Y])$$

Let's look at $L[X|Y]$ when $\mathbb{E}[X] = \mathbb{E}[Y] = 0$



$$L[X|Y] = bY = \mathrm{Proj}_{\vec{Y}} \vec{X} = \frac{\langle \vec{X}, \vec{Y}\rangle}{\|\vec{Y}\|^2} \cdot Y$$

$$= \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]} \cdot Y$$

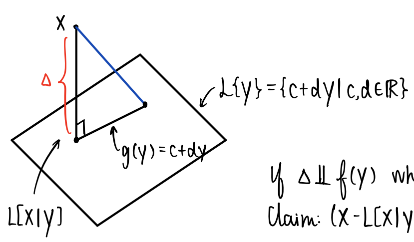
$$= \underbrace{\frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)}}_{b} \cdot Y$$

Alternatively: $(X - bY) \perp\!\!\!\perp Y \rightarrow \langle X - bY, Y\rangle = 0$

$\mathbb{E}[(X - bY)Y] = 0$

$\mathbb{E}[XY] = b\,\mathbb{E}[Y^2]$

$b = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]} = \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)}$



If $\Delta \perp\!\!\!\perp f(Y)$ where $f(Y) \in \mathcal{L}(Y)$,

Claim: $(X - L[X|Y])^2 \leq (X - g(Y))^2$

$$\|X - g(Y)\|^2 = \|X - L[X|Y] + L[X|Y] - g(Y)\|^2$$

$$= \|X - L[X|Y]\|^2 + \underbrace{\|L[X|Y] - g(Y)\|^2}_{\geq 0} + \underbrace{2\langle \overbrace{(X - L[X|Y]}^{\Delta}, (L[X|Y] - g(Y))\rangle}_{0}$$

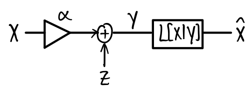
$\Rightarrow \|X - g(Y)\|^2 \geq \|X - L[X|Y]\|^2$ w/ equality iff $g(Y) = L[X|Y]$

al

Ex: $Y = \alpha X + Z$, X, Z are zero mean & indep

(sign, noise)

Find $L[X|Y]$



① Algebra:

$$L[X|Y] = \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)} \cdot Y$$

$$\mathrm{cov}(X,Y) = \mathbb{E}[XY] = \mathbb{E}[X(\alpha X + Z)]$$

$$= \alpha \mathbb{E}[X^2] + \overset{0}{\mathbb{E}[\cancel{X}Z]}$$

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] = \mathbb{E}[(\alpha X + Z)^2]$$

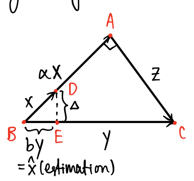
$$= \alpha^2 \mathbb{E}[X^2] + \mathbb{E}[Z^2]$$

$$\Rightarrow L[X|Y] = \frac{\alpha \mathbb{E}[X^2]}{\alpha^2 \mathbb{E}[X^2] + \mathbb{E}[Z^2]} \cdot Y$$

$$= \frac{\alpha^{-1} Y}{1 + \frac{1}{SNR}}$$

$$SNR = \frac{\alpha^2 \mathbb{E}[X^2]}{\mathbb{E}[Z^2]} = \frac{\text{signal power}}{\text{noise power}}$$

② *Geometry*: $Y = \alpha X + Z$; X, Z indep



$\alpha\vec{X} = BA$

$\vec{X} = BD$

$\Delta = DE$

$\hat{X} = bY = BE$

Find $\hat{X} = b\vec{Y} = BE$: use geometry

BDE & BAC are similar triangles

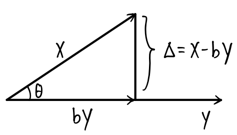
$$\frac{BE}{BD} = \frac{BA}{BC}$$

$$BE = BD\,\frac{BA}{BC}$$

$$b|\vec{Y}| = \frac{\alpha|\vec{X}| \cdot |\vec{X}|}{|\vec{Y}|}$$

$$b = \frac{\alpha|\vec{X}|^2}{|\vec{Y}|^2}$$

$$b = \frac{\alpha \mathbb{E}[X^2]}{\mathbb{E}[Y^2]} = \frac{\alpha \mathbb{E}[X^2]}{\alpha^2 \mathbb{E}[X^2] + \mathbb{E}[Z^2]}$$



$$\cos\theta = \rho = \frac{\langle X, Y\rangle}{\|X\| \cdot \|Y\|}$$

$$\mathbb{E}[\Delta^2] = \xi = \mathbb{E}[(X - \hat{X})^2] = ?$$

$$\|\Delta\|^2 = \|X\|^2 \sin^2\theta$$

$$= \|X\|^2 (1 - \cos^2\theta)$$

$$= \|X\|^2 \left[1 - \frac{\langle X, Y\rangle^2}{\|X\|^2 \cdot \|Y\|^2}\right]$$

$$= \|X\|^2 - \frac{\langle X, Y\rangle^2}{\|Y\|^2}$$

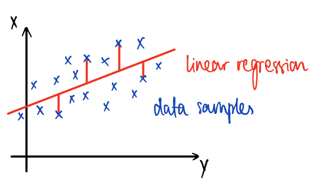
$$\mathbb{E}[\Delta^2] = \mathbb{E}[X^2] - \frac{\mathbb{E}^2[XY]}{\mathbb{E}[Y^2]}$$

$$\boxed{\xi = \mathbb{E}[\Delta^2] = Var(X) - \frac{Cov^2(X,Y)}{Var(Y)}} \quad \text{(zero mean case)}$$

seeing Y helps us reduce the error (if X, Y correlated)

Non-Bayesian View of LLSE: Linear Regression

So far, we have assumed a Bayesian framework, assuming complete knowledge of joint distr of X & Y.
Let's take a non-probabilistic (data-driven) perspective.



We have access to samples $\{(X_1, Y_1), \ldots, (X_k, Y_k)\}$.

Goal: Construct $g(Y) = a + bY$ s.t.

$$\boxed{\xi(a,b) = \frac{1}{k}\sum_{i=1}^{k} |X_i - a - bY_i|^2 \text{ is minimized}}$$

But this is identical to a Bayesian perspective where $(X, Y) \sim Uniform\{(X_i, Y_i)\}_{i=1}^{k}$

Set $\frac{\partial S}{\partial a} = 0$; $\frac{\partial S}{\partial b} = 0$

$$\Rightarrow \boxed{a + bY = E_k(X) + \frac{\mathrm{cov}_k(X,Y)}{\mathrm{Var}_k(Y)}\left(Y - E_k(Y)\right)}$$

where $E_k(X) = \frac{1}{k}\sum_{i=1}^{k} X_i$; $E_k(Y) = \frac{1}{k}\sum_{i=1}^{k} Y_i$

$$\mathrm{Var}_k(Y) = \frac{1}{k}\sum_{i=1}^{k} Y_i^2 - E_k^2(Y)$$

$$\mathrm{cov}_k(X,Y) = \frac{1}{k}\sum_{i=1}^{k} X_i Y_i - E_k(X)E_k(Y)$$

Theorem 7.3 (Walrand):

Linear Regression converges to LLSE, i.e. $LR \xrightarrow{k\to\infty} LLSE$ by SLLN.