Lecture 14: Estimation.

$$X \longrightarrow \boxed{\text{Channel}} \xrightarrow{\;y\;} \boxed{\text{Estimator}} \longrightarrow \hat{X}$$

$f_X(x)$ $\quad$ $f_{Y|X}(y|x)$

$\qquad\quad P_{Y|X}(y|x)$

Key difference from hypothesis testing: y is continious.

<u>Goal</u>: Given a probabilistic model $f_X(x)\, f_{Y|X}(y|x)$

and a cost function $c(x, \hat{x}) : \mathbb{R}^2 \longrightarrow \mathbb{R}^+$

where $c(x, \hat{x})$ measures the error we make if estimate x as $\hat{x}$.

find an estimator $g : \mathcal{Y} \longrightarrow \mathcal{X}$ s.t $\hat{x} = g(y)$

s.t. $\mathbb{E}\left[c(x, \hat{x})\right] = \mathbb{E}\left[c(x, g(y))\right]$ is minimized

<u>Ex</u>: $c(x, \hat{x}) = \begin{cases} 1 & \text{if } x \neq \hat{x} \\ 0 & \text{o/w} \end{cases}$.

$$\mathbb{E}\left[c(x, \hat{x})\right] = 1\, P\{x \neq \hat{x}\} = P_e.$$

Most popular cost function: $c(x, \hat{x}) = (x - \hat{x})^2$

$$c(\bar{x}, \hat{\bar{x}}) = \| \bar{x} - \hat{\bar{x}} \|^2 = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$$

If the cost function is squared loss this is called minimum mean squared error MMSE estimation

Minimum mean squared estimation

Find $g: \mathcal{Y} \to \mathcal{X}$ s.t $\mathbb{E}\left[(X - g(Y))^2 \right]$ is minimized

for a given model $f_X(x) \, f_{Y|X}(y|x)$ for X and Y

Simpler problem: Find a number $a \in \mathbb{R}$ s.t

$\mathbb{E}\left[(X-a)^2 \right]$ is minimized for a given $X \sim f_X(x)$

$$\mathbb{E}\left[(X - \mu + \mu - a)^2 \right] = \mathbb{E}\left[(X-\mu)^2 + 2(X-\mu)(\mu-a) + (\mu-a)^2 \right]$$

$$= \mathbb{E}\left[(X-\mu)^2 \right] + 0 + (\mu - a)^2$$

$$= Var(X) + (\mu - a)^2$$

Conclusion: $\mathbb{E}\left[(X-a)^2 \right]$ is minimized when

$$a = \mu = \mathbb{E}[X]$$

$$\mathbb{E}\left[(X - g(Y))^2 \right] = \iint (x - g(y))^2 f_{X,Y}(x,y) \, dx \, dy$$

$$= \iint (x - g(y))^2 f_{X|Y}(x|y) \, f_Y(y) \, dx \, dy$$

$$= \int \underbrace{\left(\int (x - g(y))^2 f_{X|Y}(x|y) \, dx \right)}_{= \mathbb{E}\left[(X - g(Y))^2 \mid Y = y \right]} f_Y(y) \, dy$$

$$= \int \mathbb{E}\left[(X - g(y))^2 \mid Y = y \right] f_Y(y) \, dy$$

$$\hat{x} = g(y) = \mathbb{E}[X \mid Y = y] \rightarrow \text{Conditional expectation of } X \text{ given } Y = y$$

$$= \int x \, f_{X|Y}(x|y) \, dy$$

$g(y) = \mathbb{E}[X|y] \longrightarrow$ conditional expectation of X given $y \Rightarrow$ a function of y!

<u>Conclusion:</u> The estimator function that minimizes the mean-squared error is given by

$$g(y) = \mathbb{E}[X|y]$$ the MMSE estimate.

$$\mathbb{E}[g(y)] = \mathbb{E}[\mathbb{E}[X|y]] = \int\left(\int x\, f_{X|Y}(x|y)\,dx\right) f_Y(y)\,dy$$

$$= \iint x\, f_{X|Y}(x|y)\, f_Y(y)\, dx\, dy$$

$$= \iint x\, f_{X,Y}(x,y)\, dx\, dy$$

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

$\Rightarrow$ The MMSE estimate is unbiased.

Law of iterated expectations:

$$\mathbb{E}[f(x,y)] = \mathbb{E}_y[\mathbb{E}_x[f(x,y)|y]]$$

$$= \mathbb{E}_x[\mathbb{E}_y[f(x,y)|x]]$$

$$\mathbb{E}[X] = \mathbb{E}_y[\mathbb{E}_x[X|y]]$$

Error for MMSE Estimation:

MMSE : $\mathbb{E}[(X - \mathbb{E}[X|y])^2]$

$$= \mathbb{E}_y\left(\mathbb{E}_x[(X - \mathbb{E}[X|y])^2 | y]\right)$$

$$= \int \underbrace{\mathbb{E}[(X-\mathbb{E}[X|Y])^2 | Y=y]}_{\substack{= \mathbb{E}[(X-\mathbb{E}[X|Y=y])^2|Y=y] \\ = Var(X|Y=y)}} f_Y(y)\, dy$$

$$= \int Var(X|Y=y) f_Y(y)\, dy$$

$$= \mathbb{E}[Var(X|Y)]$$

↳ conditional variance of X given Y

<u>Example:</u> X and Y are jointly Gaussian and zero-mea

$$K_{\begin{bmatrix} X \\ Y \end{bmatrix}} = \begin{bmatrix} \sigma_X^2 & Cov(X,Y) \\ Cov(X,Y) & \sigma_Y^2 \end{bmatrix}$$

We observe Y and want to estimate X.

$$\hat{X} = \mathbb{E}[X|Y] \qquad f_{X|Y}(x|y)$$

<u>Claim:</u> If X and Y are jointly Gaussian, we can always write

$$X = \alpha^* Y + W$$

for some $\alpha^* \in \mathbb{R}$ and W Gaussian and independent of Y.

<u>Proof:</u> For any α

$W = X - \alpha Y$ is always Gaussian and zero-mean

If we choose α st $\mathbb{E}[Wy] = \mathbb{E}[(X-\alpha y)y] = 0$.

$$\mathbb{E}[(X-\alpha y)y] = \mathbb{E}[Xy] - \alpha\,\mathbb{E}[yy] = 0.$$

$$\alpha^* = \frac{\mathbb{E}[Xy]}{\mathbb{E}[y^2]} = \frac{Cov(X,y)}{\sigma_y^2}.$$

then W and y are uncorrelated.

$\Rightarrow$ W and y are independent (because they are jointly Gaussian $\begin{bmatrix} W \\ y \end{bmatrix} = \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ y \end{bmatrix}$)

$$X = \alpha^* y + W$$

Conditioned on $y = y$.

$$X = \alpha^* y + W \sim N(\alpha^* y, \sigma_w^2).$$

$$\sigma_w^2 = \mathbb{E}[(X-\alpha^* y)^2] = \mathbb{E}[X^2] - 2\alpha^*\mathbb{E}[Xy] + \alpha^{*2}\mathbb{E}[y^2]$$

$$= \sigma_X^2 - 2\frac{\mathbb{E}[Xy]^2}{\sigma_y^2} + \frac{\mathbb{E}[Xy]^2}{\sigma_y^2}$$

$$= \sigma_X^2 - \frac{\mathbb{E}[Xy]^2}{\sigma_y^2}.$$

Definition: Correlation coefficient of X and y

$$\rho = \frac{Cov(X,y)}{\sigma_X \sigma_y} \qquad -1 \le \rho \le 1.$$

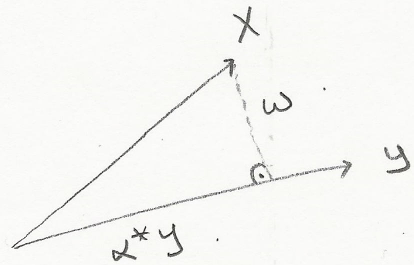
$$= \sigma_X^2 - \rho^2\sigma_X^2 = (1-\rho^2)\sigma_X^2.$$

$$\Rightarrow \mathbb{E}[X \mid Y=y] = \alpha^* y = \frac{Cov(X,Y)}{\sigma_Y^2}\, y = \rho \frac{\sigma_X}{\sigma_Y}\, y$$

$$\hat{X} = \mathbb{E}[X \mid Y] = \alpha^* Y = \rho \frac{\sigma_X}{\sigma_Y}\, Y$$

$$MSE = \mathbb{E}[Var(X \mid Y)]$$
$$= \int Var(X \mid Y=y)\, f_Y(y)\, dy = \sigma_W^2 = (1-\rho^2)\,\sigma_X^2$$

Interpretation:



$$X = \underbrace{\alpha^* Y}_{\hat{X}} + \underbrace{W}_{X-\hat{X}}$$

$$\sigma_X^2 = \alpha^{*2}\sigma_Y^2 + \sigma_W^2$$

Extension to non-zero mean

$$\tilde{X} = X - \mu_X \qquad \tilde{Y} = Y - \mu_Y$$

$$Cov(\tilde{X},\tilde{Y}) = Cov(X,Y)$$
$$Var(\tilde{X}) = Var(X)$$
$$Var(\tilde{Y}) = Var(Y)$$

$$\tilde{X} = \alpha^* \tilde{Y} + W \quad \text{where } \alpha^* = \frac{Cov(X,Y)}{Var(Y)} = \rho\frac{\sigma_X}{\sigma_Y}$$

$$X - \mu_X = \alpha^*(Y-\mu_Y) + W$$

$$X = \mu_X + \alpha^*(Y-\mu_Y) + W \quad \text{where } W \sim \mathcal{N}(0,\sigma_W^2)$$

$$\mathbb{E}[X \mid Y=y] = \mu_X + \alpha^*(y-\mu_Y)$$

$$MSE = (1-\rho^2)\,\sigma_X^2$$

HW6 pr 1 : $X \mid Y=y \sim \mathcal{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y),\ (1-\rho^2)\sigma_X^2\right)$