Note Title 4/1/2011

x term: $-2a+3b=-2$
y term: $-8b+3a=3$ $\Rightarrow b=0 \quad a=1$

Constant term: $a^2-3ab+4b^2=1$ ✓

$$f_{XY}=\frac{1}{2\pi \times c}e^{-\frac{1}{2}\left((x-1)^2-3(x-1)y+4y^2\right)}$$

$\Rightarrow$ ① $\sigma_X\sigma_Y\sqrt{1-\rho^2}=c$

② $\frac{1}{2(1-\rho^2)\sigma_X^2}=\frac{1}{2}$

③ $\frac{1}{2(1-\rho^2)\sigma_Y^2}=\frac{4}{2}$

④ $\frac{2\rho}{2(1-\rho^2)}\times\frac{1}{\sigma_X\sigma_Y}=\frac{3}{2}$

$\sigma_Y^2=4/7$

$\sigma_X^2=\frac{16}{7}$

$c=\frac{2}{\sqrt{7}}$

$\rho=\frac{3}{4}$

④^2/②③ $\Rightarrow \rho^2=\frac{9}{16} \xRightarrow{\text{by ④}} \rho=\frac{3}{4}$

substitute ρ into ② $\Rightarrow \sigma_X^2=\frac{16}{7}$

ρ into ③ $\sigma_Y^2=\frac{4}{7}$

from ① $\Rightarrow c=\frac{2}{\sqrt{7}}$

$\Rightarrow m_X=1,\ m_Y=0\quad \sigma_X^2=\frac{16}{7},\ \sigma_Y^2=\frac{4}{7}\quad c=\frac{2}{\sqrt{7}}\quad \rho=\frac{3}{4}$

$Cov(X,Y)=\frac{3}{4}\times\frac{8}{7}$

Note Title 4/4/2011

* 2-dim Joint Gsn R.V. (X,Y)

S_{XY} : {all real 2-dim vectors}.

$$f_{XY}(x,y) = \frac{1}{2\pi\,\sigma_X\sigma_Y\left(\sqrt{1-\rho^2}\right)} \times e^{-\frac{\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-m_X}{\sigma_X}\right)\left(\frac{y-m_Y}{\sigma_Y}\right) + \left(\frac{y-m_Y}{\sigma_Y}\right)^2}{2(1-\rho^2)}}$$

* Properties of joint Gsn R.V.s.

① $E(X) = m_X$, $Var(X) = \sigma_X^2$

$Cov(X,Y) = \rho \cdot \sigma_X \cdot \sigma_Y$

② The marginal distribution of X is Gsn, The marginal distribution of Y

is Gsn. Moreover, any linear combination of X & Y is (joint) Gsn

ex: $Z = 3X + 4Y$ is Gsn
$W = 2X - Y$ is Gsn
& (Z, W) are $\boxed{\text{joint Gsn}}$

Ex: X, Y are joint Gsn
with $m_X = 1$, $m_Y = 0$, $\sigma_X = 1$, $\sigma_Y = 2$
$\rho = -0.5$

Q: marginal pdf of $Y = ?$

Ans: by ② $\Rightarrow f_Y(y) = \frac{1}{\sqrt{2\pi 2^2}} e^{-\frac{(y-0)^2}{2 \times 2^2}}$

Q: $Z = 3X + 4Y$. Find m_Z, σ_Z.

Ans: $m_Z = 3m_X + 4m_Y = 3 \times 1 + 4 \times 0 = 3$

$\sigma_Z^2 = 3^2 Var(X) + 2 \times 3 \times 4 \, Cov(X, Y) + 4^2 Var(Y)$
$= 9 \times 1 + 24 \times (-0.5 \times 1 \times 2) + 16 \times 4$
$= 49$

* Gsn $\rightarrow$ $\boxed{\text{Linear}}$ $\xrightarrow{\text{Gsn}}$

* The benefit of working on a Gsn is that we only need to worry about the mean, variance, covariance of its input / output

We don't need to worry about $f_{XY}(x, y)$

Properties

③ If X and Y both are Gsn
& X, Y are independent

$\Rightarrow (X, Y)$ are $\boxed{\text{joint Gsn}}$.

Exercise: Find a joint distribution (X, Y) s.t
X & Y are both Gsn but (X, Y) is not joint Gsn.

④ Generally independent $\Longrightarrow$ uncorrelated
$\not\Longleftarrow$ Not vice versa.

but if X & Y are joint Gsn,
then independent $\Longleftrightarrow$ uncorrelated

pf: Look at the Gsn joint pdf formula.

Ex: X & Y are standard Gsn.
& X & Y are independent.

$Z = X + Y$
$W = X - Y$

Q: Are Z, W joint Gsn.

Ans: $\because (X, Y)$ are joint Gsn
$\therefore (Z, W)$ the linear combination of X and Y
are joint Gsn.

Q: m_Z, σ_Z, m_W, σ_W, $Cov(Z,W) = ?$

Ans: $m_Z = m_X + m_Y = 0$

$m_W = m_X - m_Y = 0$

$$\sigma_Z^2 = \underbrace{\sigma_X^2}_{1} + \underbrace{2\,Cov(X,Y)}_{0 \;\because\; \rho = 0 \text{ by independence of } X, Y} + \underbrace{\sigma_Y^2}_{1}$$

$= 2$,

$$\sigma_W^2 = \sigma_X^2 - 2\,Cov(X,Y) + \sigma_Y^2$$

$= 2$

$$Cov(Z,W) = E((Z-0)(W-0))$$
$$= E(Z \cdot W)$$
$$= E((X+Y)(X-Y))$$
$$= E(X^2 - Y^2)$$
$$= E(X^2) - E(Y^2)$$
$$= (1 + 0^2) - (1 + 0^2)$$
$$= 0$$

Q: Are Z, W independent?

Ans: Yes. $\because$ $Cov(Z, W) = 0$

Q: $f_{ZW}(z, w) = ?$

$$\text{Ans: } \frac{1}{2\pi\sqrt{2\times 2}} e^{-\frac{\left(\frac{z-0}{\sqrt{2}}\right)^2 - 0(\ )(\ ) + \left(\frac{w-0}{\sqrt{2}}\right)^2}{2\times(1-0^2)}}$$

$$= \frac{1}{2\pi\sqrt{2\times 2}} e^{-\frac{z^2+w^2}{2\times 2}} \quad \#$$

Property ⑤ if X, Y are joint Gsn. then $P(X \mid Y=y)$, the conditional distribution of X is also Gsn with mean

$$m_X + \rho \times \frac{\sigma_X}{\sigma_Y} \times (y - m_Y)$$

variance

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

see p. 281 for derivation

* Detection & estimation $X \rightarrow \square \rightarrow Y$

The original quantity X is unknown, we only observe Y. jointly X & Y are randomly distributed. Our goal is to derive the information of X from the observation Y.

Ex: X: Signals at the base station

Y: Signals received by the cellular phone.

Ex: X: Waveform in a concert.

Y: Recorded MP3 signals.

Ex: X: The # of users login to a web server

Y: The download speed of from my dormitory

Ex: X: The exact location of a missile

Y: The radar output.

* Detection & Estimation $X \to \boxed{} \xrightarrow{Y}$

There are many different schems for detection & estimation with different performance complexity tradeoff.

Scheme 1: Maximum a posterior prob. (MAP) detector.

We first observe $Y = y_0$.
Find the x with the largest condition prob.

prob $P(X=x \mid Y=y_0)$

similar to
Ex: HW8 Q8 Prob 6.68

$X \backslash Y$	-1	0	1
-1	$\frac{1}{12}$	$\frac{1}{6}$	0
0	0	0	$1/3$
1	$\frac{1}{4}$	$1/6$	0

Q: Find the MAP detector given $Y = y_0$.

Ans: Given $Y=1$ $P(X=x \mid Y=1)$

$$= \begin{cases} 0 & \text{if } x=-1 \\ 1 & \text{if } x=0 \\ 0 & \text{if } x=1 \end{cases}$$

Given $Y=0$, $P(X=x \mid Y=0)$

$$= \begin{cases} \frac{1}{2} & \text{if } x=-1 \\ 0 & \text{if } x=0 \\ \frac{1}{2} & \text{if } x=1 \end{cases}$$

$Y=-1$ $\quad P(X=x \mid Y=-1)$

$$= \begin{cases} \frac{1}{4} & \text{if } x=-1 \\ 0 & \text{if } x=0 \\ \frac{3}{4} & \text{if } x=1 \end{cases}$$

$$\Rightarrow MAP(y) = \begin{cases} 1 & \text{if } y=-1 \\ \text{either } -1 \text{ or } 1 \text{ is fine} & \text{if } y=0 \\ 0 & \text{if } y=1 \end{cases}$$

☆ MAP has very strong performance (optimal)

as we always choose the most probable X given the observation $Y=y$

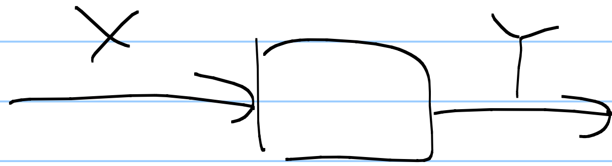
☆ The drawback of MAP is ~~its~~ complexity

Both Finding conditional prob & finding the maximum are difficult to implement.

$\boxed{\text{Start from } P_X \to P_X \cdot P_{Y|X} \to P_{X|Y}}$

195 * Scheme 2: Maximum Likelihood (ML) detector.

Recall that we have observed $Y = y_0$



and we are interested in infering X.

* Define the likelihood of $Y = y_0$ as $f_{\text{Likelihood}}(x) = P(Y = y_0 \mid X = x)$

The Maximum Likelihood (ML) detector thus outputs x that has the largest

$$f_{\text{Likelihood}}(x) = P(Y = y_0 \mid X = x)$$

$\boxed{\text{Comparison}}$

Maximum Aposteriori Prob (MAP) detector outputs x that has the largest

$$P(X = x \mid Y = y_0)$$

Q: When do we use ML instead of MAP?

Ans: Sometimes we do not have the original marginal P_X. In this case, we just assume

$\boxed{P_X \text{ is uniform}}$

then $P(X=x \mid Y=y_0)$

$$= \frac{P(X=x, Y=y_0)}{P(Y=y_0)}$$

$$= \frac{P(Y=y_0 \mid X=x) \times \boxed{P(X=x)}}{\boxed{P(Y=y_0)}}$$

Do not depend on x

maximizing$_x$ $P(X=x \mid Y=y_0)$ a posterior prob.

$= $ maximizing $P(Y=y_0 \mid X=x)$

the likelihood

* ML can be viewed as a special case of MAP when P_X (the prior) is uniform.

* Example: Conditional prob.

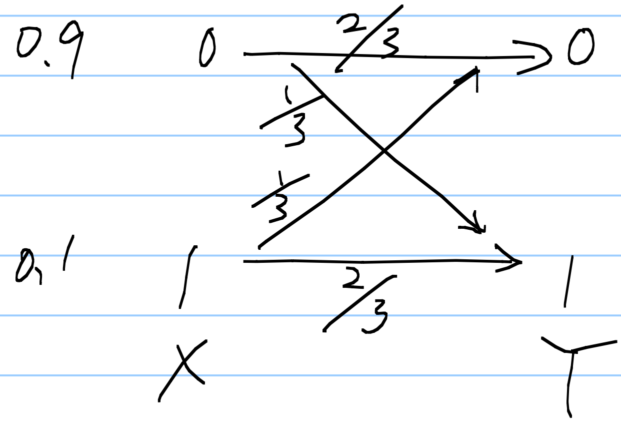
$P(Y=0 \mid X=0) = \frac{2}{3}$ $\quad P(X=0) = 0.9$

$P(Y=1 \mid X=0) = \frac{1}{3}$ $\quad P(X=1) = 0.1$

$P(Y=0 \mid X=1) = \frac{1}{3}$

$P(Y=1 \mid X=1) = \frac{2}{3}$

This is sometimes called the <u>binary symmetric channel</u>.



Find $MAP(y)$; $ML(y)$

Ans: $ML(0)$: Comparing likelihood

$$P(Y=0 \mid X=0) = \frac{2}{3} \quad \checkmark$$

$$P(Y=0 \mid X=1) = \frac{1}{3}$$

$ML(1)$ ... Comparing

$$P(Y=1 \mid X=0) = \frac{1}{3}$$

$$P(Y=1 \mid X=1) = \frac{2}{3} \quad \checkmark$$

$$\Rightarrow ML(y) = \begin{cases} 0 & \text{if } y=0 \\ 1 & \text{if } y=1 \end{cases}$$

$MAP(0)$: Comparing the conditional prob (posterior)

$$P(X=0 \mid Y=0) = \frac{0.9 \times \frac{2}{3}}{0.9 \times \frac{2}{3} + 0.1 \times \frac{1}{3}} = \frac{18}{19}$$

$$P(X=1 \mid Y=0) = \frac{0.1 \times \frac{1}{3}}{0.9 \times \frac{2}{3} + 0.1 \times \frac{1}{3}} = \frac{1}{19}$$

MAP(1) : comparing

$$P(X=0 \mid Y=1) = \frac{0.9 \times \frac{1}{3}}{0.9 \times \frac{1}{3} + 0.1 \times \frac{2}{3}}$$

$$= \frac{9}{11}$$

$$P(X=1 \mid Y=1) = \frac{0.1 \times \frac{2}{3}}{0.9 \times \frac{1}{3} + 0.1 \times \frac{2}{3}}$$

$$= \frac{2}{11}$$

$$\Rightarrow MAP(y) = \begin{cases} 0 & \text{if } y=0 \\ 0 & \text{if } y=1 \end{cases}$$

$\because$ $P(X=0)=0.9$ is much more possible than $P(X=1)$.

MAP takes that into account while ML does not.

MAP detector has optimal performance, but higher Complexity (finding $P_{X|Y} = \frac{P_{Y|X} P_X}{P_Y}$)

ML detector has slightly poorer performance & less complexity, (working on $P_{Y|X}$)

* MMSE estimater

Both ML, MAP detectors take an all-or-nothing approach. It either misses or it hits. There is no middle point.

For comparison, MMSE estimater find an estimation that is the closest to all possible outcomes.

Example: Five houses on a street

Q1: Determine a location $\hat{x}$ that has the smallest

$$\sum_{k=1}^{5} (x_k - \hat{x})^2$$

$\boxed{\text{called the \underline{objective} function}}$

Q2: We know that each house has a prob p_k to be on fire. We need to find a location $\hat{x}$ s.t.

$$\sum_{k=1}^{5} (x_k - \hat{x})^2 \cdot p_k$$

or equivalent $E((X - \hat{x})^2)$ is minimized

Minimal Mean Square Error

Error $(X - \hat{x})$ Squar ^{2}. Mean E, Minimal

* The constant $\hat{x}$ that minimizes $E((X-\hat{x})^2)$ is called the MMSE estimator.

The resulting "distance" $E((X-\hat{x})^2)$ is called MMSE.

* The MMSE estimator is computed by

$$\hat{x}_{MMSE} = E(X)$$

pf: $\frac{d}{d\hat{x}} E((X-\hat{x})^2) = 0$

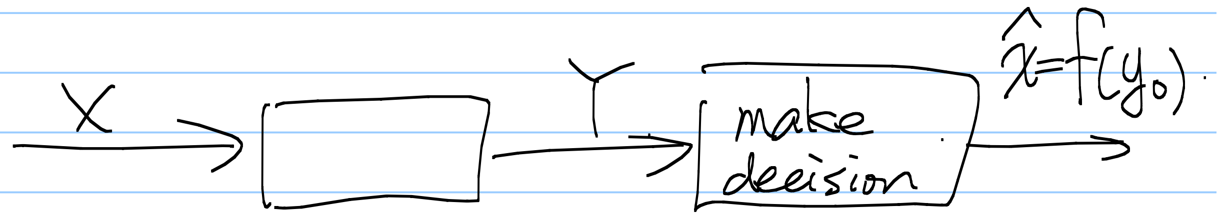
$$= E\left(\frac{d}{d\hat{x}}(X-\hat{x})^2\right) = 0$$

$$= E(2(X-\hat{x})\cdot(-1)) = 0$$

$$= -2E(X) + 2\hat{x}$$

$$\hat{x} = E(X)$$

Recall that oftentimes we can observe Y before we have to make a decision on X. That is



Question: Observing $Y = y_0$, how to choose the MMSE estimator?

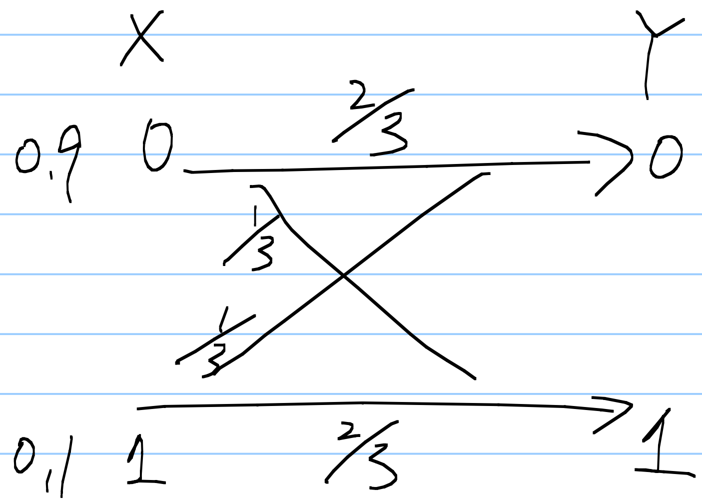
Answer: $\hat{X}_{MMSE} = E(X \mid Y = y_0)$

Observing $Y = y_0$

$\hat{X}_{MMSE} = E(X \mid Y = y_0)$ the "center" of the conditional prob is the closest estimation one can have under $Y = y_0$

Note: $\hat{X}_{MMSE}(y)$, $MAP(y)$, $ML(y)$ are all functions of y.

Continue from our example



$P(X=0 \mid Y=0) = \frac{18}{19}$

$P(X=1 \mid Y=0) = \frac{1}{19}$

$P(X=0 \mid Y=1) = \frac{9}{11}$

$P(X=1 \mid Y=1) = \frac{2}{11}$

Q: What is the MMSE estimator $\hat{x}_{MMSE}(y)$

Ans:

$$\hat{x}_{MMSE}(0) = E(X \mid Y=0) = 0 \times \frac{18}{19} + 1 \times \frac{1}{19} = \frac{1}{19}$$

$$\hat{x}_{MMSE}(1) = E(X \mid Y=1) = 0 \times \frac{9}{11} + 1 \times \frac{2}{11} = \frac{2}{11}$$

Note that $\hat{x}_{MMSE}(y)$ does not hit any of the outcome. But it is closest to all outcomes.

Drawback: Still need to compute $P_{X|Y}$, which is hard to implement

Summary

Note Title 4/12/2011

* Detection & estimation

Prob modeling

P_X $P_{Y|X}$

Goal: design good function st. $\hat{x}(y_0)$ is a "good substitute" of the true X

Scheme 1. MAP detector → maximize the hit prob.

Output x that maximizes

$$P(X=x \mid Y=y_0)$$

$\hat{x}_{MAP}$ is a function $\hat{x}_{MAP}(y_0)$.

Scheme 2 ML detector → a simpler version of MAP by assuming uniform X

Output x that maximizes

$$P(Y=y_0 \mid X=x)$$

$\hat{x}_{ML}$ is a function of $\hat{x}_{ML}(y_0)$

Scheme 3: Minimal Mean Square Error estimator → Never hits X exactly. But is always close to X

minimize $E((X-\hat{x})^2)$

$\hat{x}_{MMSE}$ is a function $\hat{x}_{MMSE}(y)$

* Linear MMSE estimator

Note Title 4/12/2011

$\hat{x}_{Lin,MMSE}(y) = ay + b$ a linear function of y that minimizes

$$E\left(\left(X - (aY + b)\right)^2 \right)$$

the minimizing a, b are

$$a^* = \frac{\rho_{XY}\,\sigma_X}{\sigma_Y}$$

$$b^* = -\frac{\rho_{XY}\,\sigma_X\,m_Y}{\sigma_Y} + m_X$$

$$a^* y + b^* = \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - m_Y) + m_X$$

Since $m_X, m_Y, \sigma_X, \sigma_Y, \rho_{XY}$ are easy to compute, we can obtain a^*, b^* very efficiently.

Then every time we observe $Y = y_0$

$$\hat{x}_{Lin,MMSE}(y_0) = a^* y_0 + b^*.$$