L02

Monday, August 15, 2022 1:31 PM

Inference Prob:

$$\boxed{P_X} \xrightarrow{X} \boxed{P_{Y|X}} \xrightarrow{Y} \boxed{f(Y)} \xrightarrow{\hat{X}}$$

inference

$$\hat{X}_{MAP}(y) = \underset{x}{\operatorname{argmax}} \; P_{X|Y}(x|y)$$

$$\hat{X}_{ML}(y) = \underset{x}{\operatorname{argmax}} \; P_{Y|X}(y|x)$$

A table-based example. Instead of P_X, $P_{Y|X}$, we are given the joint prob.

The joint prob of X & Y is

$Y \backslash X$	0	1
0	$\frac{1}{9}$	[illegible]/24
1	$\frac{1}{18}$	$\frac{1}{8}$
2	$\frac{1}{6}$	$\frac{1}{8}$
3	1/[illegible]	5/[illegible]

2	6	8
3	$\frac{1}{6}$	$\frac{5}{24}$

Q: Find MAP(y).

Ans: Solution 1: Compute the posterior prob as before. Find $\hat{X}_{MAP}(y)$ then.

Solution 2: For each y value (each row), the $\hat{X}_{MAP}(y)$ simply selects the block with relatively large prob.

$$\hat{X}_{MAP}(y) = \begin{cases} 0 & \text{if } y=0 \\ 1 & \text{if } y=1 \\ 0 & y=2 \\ 1 & y=3 \end{cases}$$

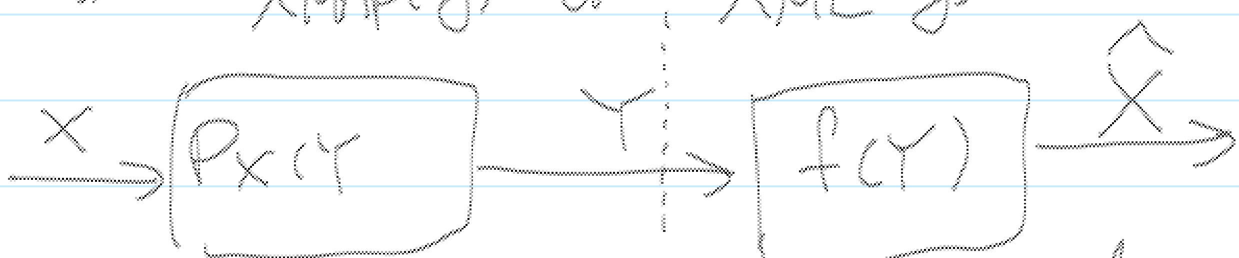
Summary:

① A detector is a function $f(y)$, where

y is the observed value (can be scalar/vector)

② $\hat{X}_{MAP}(y)$ & $\hat{X}_{ML}(y)$ are special detectors/functions of y that are developed by the knowledge about the underlying joint prob P_{XY}. (or developed by the "assumed" joint prob P_{XY}.)

Q: How to analyze the performance of a given detector $f(y)$ or the $\hat{X}_{MAP}(y)$ or $\hat{X}_{ML}(y)$?



good or bad

* Analysis of the error prob of a given detector $f(y)$?

Method 1: Conditioning on the true X value

$$P(f(Y) \neq X \mid X=0)$$
$$= P(f(Y) = 1 \mid X=0) \text{ : The false alarm prob.}$$

$$P(f(Y) \neq X \mid X=1)$$
$$= P(f(Y) = 0 \mid X=1) \text{ : The misdetection prob.}$$

Revisit the BSC example

Let us compute the <u>false alarm</u> prob.

Case 1: $p < \frac{1}{3}$

$$P(\hat{X}_{MAP}(Y) = 1 \mid X=0)$$

misdetection prob.

$P(\hat{X}_{MAP}(Y) = 1 \mid X=0)$
$= P(Y=1 \mid X=0) = \boxed{p.}$

Case 2: $\frac{1}{3} < p < \frac{2}{3}$

$P(\hat{X}_{MAP}(Y) = 1 \mid X=0)$
$= P(Y = 0 \text{ or } 1 \mid X=0) = \boxed{1}$

Case 3: $\frac{2}{3} < p$

$P(\hat{X}_{MAP}(Y) = 1 \mid X=0)$
$= P(Y=0 \mid X=0) = \boxed{1-p}$ $\times\!\!\times$

...ection p...

$P(Y=0 \mid X=1) = \boxed{p}$

$P(\text{empty set} \mid X=1) = \boxed{0}$

$P(Y=1 \mid X=1) = \boxed{1-p}$

We can compute the misdection prob similarly.

Another example:

$P_{Y|X}(\cdot \mid 0) \sim N(1, \sigma^2)$ $\quad P_X(0) = \frac{1}{3}$

$P_{Y|X}(\cdot \mid 1) \sim N(-1, \sigma^2)$ $\quad P_X(1) = \frac{2}{3}$

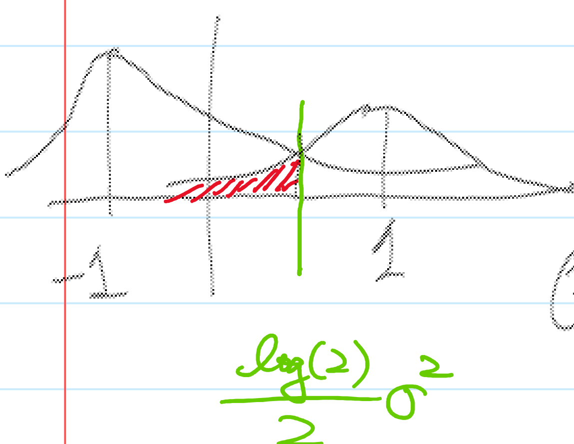
Find $\hat{X}_{MAP}(y)$ & the false-alarm prob.

Ans: Use the likelihood ratio test.

$$\hat{X}_{MAP}(y) = \begin{cases} 0 & \\ 0 \text{ or } 1 & \text{if } \dfrac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-1)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+1)^2}{2\sigma^2}}} \ \begin{matrix} > \\ = \\ < \end{matrix} \ \dfrac{\frac{2}{3}}{\frac{1}{3}} \\ 1 & \end{cases}$$

$$= \begin{cases} 0 & \\ 0 \text{ or } 1 & \text{if } y \ \begin{matrix} > \\ = \\ < \end{matrix} \ \frac{\log(2)}{2}\sigma^2 \\ 1 & \end{cases}$$

False alarm prob: $(X=0,\ \hat{X}_{MAP}(Y)=1)$



$$Q\left(\frac{1-\frac{\log(2)}{2}\sigma^2}{\sigma}\right)$$

$$Q(t) = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$$

Method 2: The overall/average error prob.

(Not just false alarm or misdetection)

$$P(f(Y) \neq X) = P(X=0)\,P(f(Y)=1 \mid X=0) + P(X=1)\,P(f(Y)=0 \mid X=1)$$

$$= P((X,Y)=(x,y) \text{ s.t. } f(y) \neq x)$$

Revisit the table-based example

Q: $P(\hat{X}_{MAP}(Y) \neq X) = ?$

Ans: It is simply the sum of the prob of those not chosen blocks.

$Y \backslash X$	0	1
0	0	$\frac{1}{24}$
1	$\frac{1}{18}$	$\frac{1}{8}$
2	$\frac{1}{6}$	$\frac{1}{8}$
3	$\frac{1}{6}$	$\frac{5}{24}$

$$P(\hat{X}_{MAP}(Y) \neq X)$$

$$= \frac{1}{24} + \frac{1}{18} + \frac{1}{8} + \frac{1}{6}$$

It seems natural for a detector to choose $\underset{x}{\arg\max}\, P_{X|Y}(x|y)$. But is it optimal? In what sense.

Theorem: $\hat{X}_{MAP}(\cdot)$ minimizes $P(f(Y) \neq X)$

Theorem: $\hat{X}_{MAP}(\cdot)$ minimizes $P(f(Y) \neq X)$

I.e. among all possible detectors, $\hat{X}_{MAP}(Y)$ has the smallest overall error prob.

Pf: We prove it by considering problems that are converted to its table form.

For any y, $\hat{X}_{MAP}(y)$ selects the block with the largest $P_{XY}(x, y)$

$\Leftrightarrow$ for any y, $\hat{X}_{MAP}(y)$ minimizes

$$\sum_{x:\, x \neq \hat{X}_{MAP}(y)} P_{XY}(x, y)$$

Summing this prob for all y values gives the overall error prob.

$\Rightarrow$ $\hat{X}_{MAP}(\cdot)$ minimizes the overall error prob.

Theorem: $P(\hat{X}_{MAP}(Y) \neq X) \leq \frac{|S_X|-1}{|S_X|}$

proof 1: Compare it to a blind rule

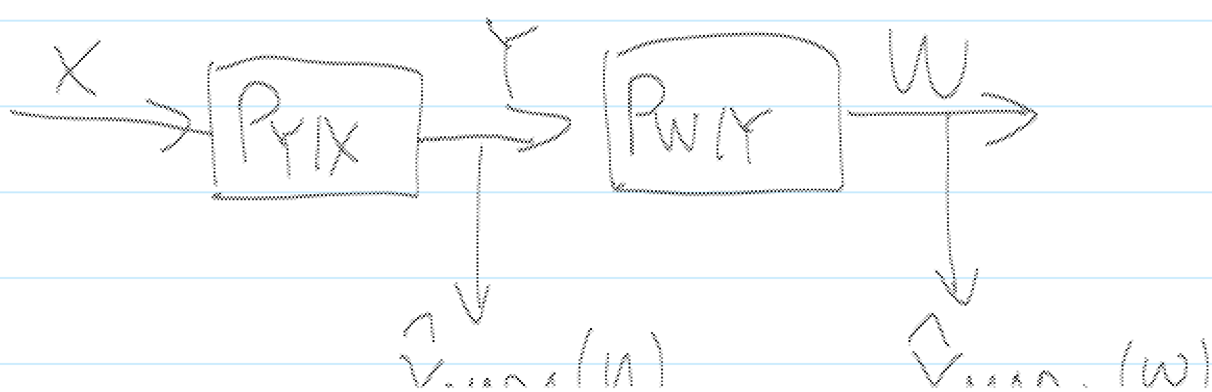
proof 2: $\max_x P_{XY}(x,y) \geq \frac{1}{|S_X|} \sum_x P_{XY}(x,y)$

$$\Rightarrow \sum_{x : x \neq \hat{X}_{MAP}(y)} P_{XY}(x,y) \leq \frac{|S_X|-1}{|S_X|} \sum_x P_{XY}(x,y)$$

Summing over all y's

$$\Rightarrow P(\hat{X}_{MAP}(Y) \neq X) \leq \frac{|S_X|-1}{|S_X|} \cdot 1$$

* The error prob of degraded channels:



$\check{\hat{X}}_{MAP,1}(y) \qquad \check{\hat{X}}_{MAP,2}(w)$

Theorem: $P(\hat{X}_{MAP,1}(Y) \neq X) \leq P(\hat{X}_{MAP,2}(W) \neq X)$

proof 1: Compare it to $f(y) = \hat{X}_{MAP,2}(\tilde{W}|_y)$

By throwing a dice to generate its own $\tilde{W}$

pf 2: The table method.

Hint 1: If for any $y_1 \neq y_2$, w,

$$P(W=w \mid Y=y_1) \cdot P(W=w \mid Y=y_2) = 0$$

That is, differen y_1 and y_2 will result in disjoint set of W_1 and W_2

then $\hat{X}_{MAP,1}(Y) = \hat{X}_{MAP,2}(W)$

Hint 2:

consider

$\xrightarrow{Y} \boxed{\text{many-to-one function}} \xrightarrow{W}$

$$\longrightarrow \boxed{\text{many-to-one function}} \xrightarrow{W}$$

& use the table-method to compare

$$P(\hat{X}_{MAP,1}(Y) \neq X) \leq P(\hat{X}_{MAP,2}(W) \neq X)$$