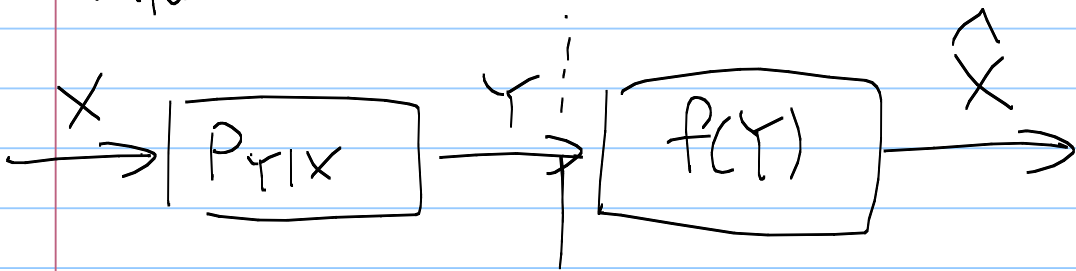
Lecture 02

1/11/2012

Inference Problem



$\hat{X}_{MAP}(y)$ & $\hat{X}_{ML}(y)$ are guided by the underlying prob distribution

$$\hat{X}_{MAP}(y) = \underset{x}{\arg\max}\, P_{X|Y}(x|y)$$

$$\hat{X}_{ML}(y) = \underset{x}{\arg\max}\, P_{Y|X}(y|x)$$

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	$\frac{1}{9}$	$\frac{1}{24}$
1	$\frac{1}{18}$	$\frac{1}{8}$
2	$\frac{1}{6}$	$\frac{1}{8}$
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.

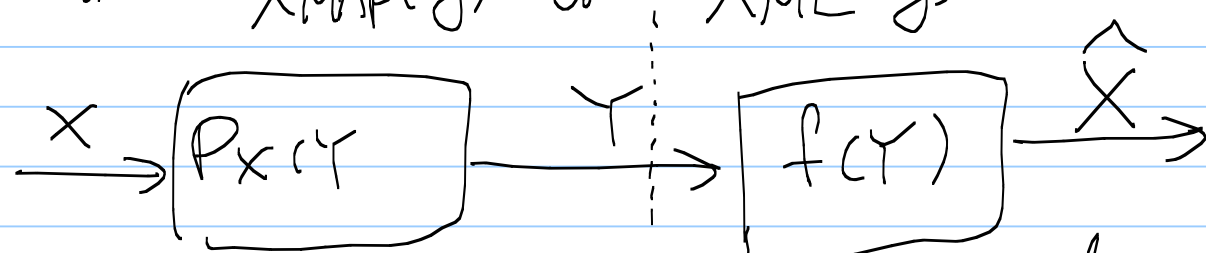
$$\hat{X}_{MAP}(y) = \begin{cases} 0 & \text{if } y=0 \\ 1 & \text{if } y=1 \\ 0 & y=2 \\ 1 & y=3 \end{cases}$$

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

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

$$P(f(Y) \neq X \mid X=1)$$
$$= P(f(Y)=0 \mid X=1)$$: The misdection prob.

Revisit the BSC example

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

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

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

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

misdection prob.

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

$$P(\phi \mid X=1) = 0$$

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

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

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

$= P(Y = 0 \mid X = 0) = 1 - p$ ※

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.

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

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

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.

$$P(f(Y)\neq X)=P(X=0)\,P(f(Y)=1\mid X=0)$$
$$+P(X=1)\,P(f(Y)=0\mid X=1)$$
$$=P((X,Y)=(x,y)\text{ s.t. } f(y)\neq x)$$

Revisit the table-based example

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

Ans: It is simply the sum of the prob of those <u>not chosen blocks</u>.

$Y \backslash X$	0	1
0	$\frac{1}{9}$	$\frac{1}{24}$
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 $\arg\max_{x} P_{X|Y}(x|y)$. But is it optimal? In what sense.

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 <u>with the largest $P_{XY}(x, y)$</u>

$\Leftrightarrow$ 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

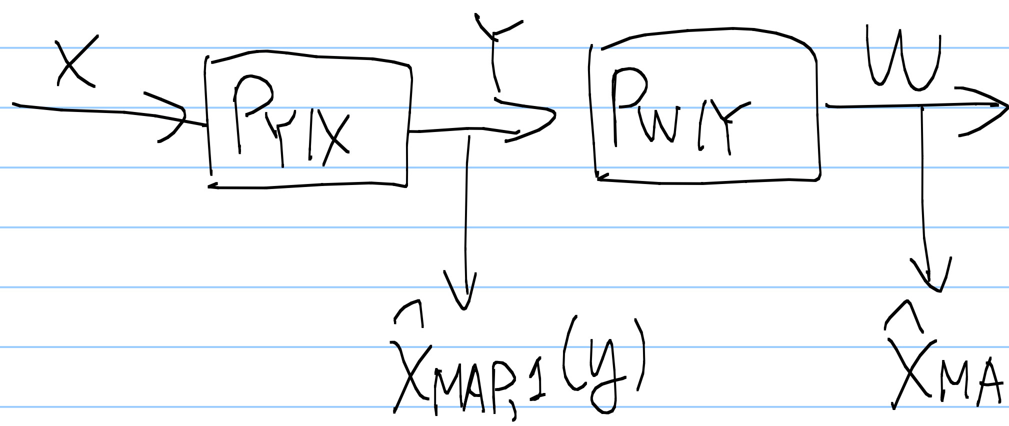
proof 2: $\max\limits_{x} P_{XY}(x,y) \geq \frac{1}{|S_X|} \sum\limits_{x} P_{XY}(x,y)$

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

Summing over all y's

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

* The error prob of degraded channels:



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 any Y will lead to a distinct (set of) W, then

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

Hint 2:

consider

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

& use the table-method to compare

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

* $\boxed{\text{Chernoff Bound}}$

Consider the log likelihood ratio test

$$\hat{X} = \begin{cases} 0 & \\ 0 \text{ or } 1 & \text{if } \log \dfrac{P_{Y|X}(y|0)}{P_{Y|X}(y|1)} \triangleq T(y) \gtreqless \tau \\ 1 & \end{cases}$$

The false alarm prob is then

$$P(T(Y) < \tau \mid X=0) \overset{\text{shorthand}}{=\!=} P_0(T(Y) < \tau)$$

Misdetection prob.

$$P_1(T(Y) > \tau)$$

Many times they are hard to compute.

$\Rightarrow$ Find a bound instead.

$\boxed{\text{Markov Inequality}}$

Assuming $P(X<0)=0$

$$\Rightarrow P(X \geq d) \leq \frac{E(X)}{d}$$

$\boxed{\text{Chernoff Bound}}$

For general X

$P(X \geq d)$

$\leq P(sX \geq sd)$ for any $s \geq 0$

$= P(e^{sX} \geq e^{sd}) \leq \dfrac{E(e^{sX})}{e^{sd}}$ for any $s > 0$

$$\leq \min_{s \geq 0} \frac{E(e^{sX})}{e^{sd}}$$

For i.i.d. $X_1, \cdots, X_n$.

Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ be the sample mean

$$P(\bar{X} \geq d) = P\left(\sum_{i=1}^{n} X_i \geq nd\right)$$

$$\leq \min_{s \geq 0} \frac{E(e^{s \sum X_i})}{e^{snd}}$$

indep.

$$= \min_{s \geq 0} \frac{\prod_{i=1}^{n} E(e^{sX_i})}{e^{snd}}$$

$$= \min_{s \geq 0} \frac{(E(e^{sX}))^n}{e^{snd}} = \left(\min_{s \geq 0} \frac{E(e^{sX})}{e^{sd}}\right)^n$$

$\Rightarrow$ For a sample mean $\overline{X}$

$P(\overline{X} \geq d) \leq (\text{The CB for each } X)^n$

for any n. (exponential decay)

Moreover, the Chernoff bound is asymptotically tight (under general conditions)

I.e. If $E(X) < d$ then $\forall \varepsilon > 0, \exists n_0$ such that

$$P(\overline{X} \geq d) \geq \left(\min_{s \geq 0} \frac{E(e^{sX})}{e^{sd}} - \varepsilon \right)^n$$

$\forall n \geq n_0$

Example: X_i = i.i.d. Bernoulli R.V.

$X_i \in_R \{0, 1\}$.

$$\overline{X}_{1000} = \frac{1}{1000} \sum_{i=1}^{1000} X_i$$

Q: $P(\overline{X}_{1000} \geq 0.75)$

Solution 1: Binomial distribution

Solution 2: Gaussian approximation (by the central limit theorem)

Solution 3: By Chernoff bound, $d = 0.75$

$$\min_{s \geq 0} \frac{E(e^{sX})}{e^{s \cdot 0.75}} = \min_{s \geq 0} \frac{\frac{1}{2} + \frac{1}{2} e^{s}}{e^{s \times 0.75}}$$

$$s^* = \log \frac{0.75}{0.25} \qquad = 2 \times 3^{-3/4} \simeq 0.8774$$

$$\Rightarrow P(\overline{X}_{1000} \geq 0.75) \leq \left(2 \times 3^{-3/4}\right)^{1000}$$