Lecture 17

(7 October 2024)

The Continuous Bayes' Rule

$$X \sim f_X \longrightarrow \boxed{f_{Y|X}} \longrightarrow Y$$

unknown RV with known PDF

Observed RV

Goal: Infer about X.

The information about X provided by the event $\{Y=y\}$ is captured by the conditional PDF $f_{X|Y}$.

$$f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y)$$

$$\Rightarrow f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{t=-\infty}^{\infty} f_X(t) f_{Y|X}(y|t)\,dt}.$$

Inference about a Discrete RV from a Continuous RV

For an event A and a continuous RV Y we first define

$$P(A \mid Y=y) = \lim_{\Delta y \to 0} P(A \mid y < Y \le y + \Delta y)$$

$$= \lim_{\Delta y \to 0} \frac{P(A,\ y < Y \le y + \Delta y)}{P(y < Y \le y + \Delta y)}$$

$$= \lim_{\Delta y \to 0} \frac{P(A)\, P(y < Y \le y + \Delta y \mid A)}{F_Y(y+\Delta y) - F_Y(y)}$$

$$= \lim_{\Delta y \to 0} \frac{P(A)\left(F_{Y|A}(y+\Delta y) - F_{Y|A}(y)\right)/\Delta y}{\left(F_Y(y+\Delta y) - F_Y(y)\right)/\Delta y}$$

$$= \frac{P(A)\, f_{Y|A}(y)}{f_Y(y)}.$$

$$\therefore P(A \mid Y = y) = \frac{P(A)\, f_{Y|A}(y)}{f_Y(y)}.$$

$$= \frac{P(A)\, f_{Y|A}(y)}{P(A)\, f_{Y|A}(y) + P(A^c)\, f_{Y|A^c}(y)},$$

Let X be a discrete RV and Y be a continuous RV.

$$X \sim p_X \longrightarrow \boxed{f_{Y|X}} \longrightarrow Y$$

$$X \in \{0, 1, 2, \ldots, M-1\}.$$

$$P(X=x \mid Y=y) = \frac{P_X(x)\, f_{Y|X}(y|x)}{\sum_{x'=0}^{M-1} P_X(x')\, f_{Y|X}(y|x')},$$

<u>Goal:</u> To estimate the hypothesis x that lead to an observation y.

A ~~test~~ $\hat{x} : \mathcal{Y} \rightarrow \mathcal{X}$ is a decision rule or a deterministic function of the observation y.

$P_X(x)$ is called 'a priori probability'.

$P_{X|Y}(x|y)$ is the probability that hypothesis x is correct on observing y.

$P_{X|Y}(x|y)$ is called 'a posteriori probability'.

Consider the decision rule that maximizes this a posteriori probability.

$$\hat{x}_{MAP}(y) = \arg\max_{x \in \{0,1,\dots,M-1\}} p_{X|Y}(x|y).$$

Maximum a posteriori probability (MAP) rule

when multiple hypothesis achieve the maximum, we choose the largest maximizing x.

For any test A, $p_{X|Y}(\hat{x}_A(y)|y)$ is the probability that $\hat{x}_A(y)$ is the correct decision when test A is used on observation y. We have

$$p_{X|Y}(\hat{x}_{MAP}(y)|y) \geq p_{X|Y}(\hat{x}_A(y)|y),$$

for all A and y.

The probability of correctness for a given test A is

$$p(\hat{x}_A(y) = x).$$

Theorem: The MAP rule maximizes the probability of correct decision conditional on each observation y. It also maximizes the overall probability of correct decision defined above.

Proof: For $x \in \{0, 1, \ldots, M-1\}$ let $A_x = \{y : \hat{x}_A(y) = x\}$ be the set of all observations y that test A maps to hypothesis x.

$$P(\hat{x}_A(Y) = X)$$

$$= \sum_{x=0}^{M-1} P(\hat{x}_A(Y) = x \mid X = x)\, P_X(x)$$

$$= \sum_{x=0}^{M-1} P(Y \in A_x \mid X = x)\, P_X(x)$$

$$= \sum_{x=0}^{M-1} \int_{y \in A_x} f_{Y|X}(y|x)\, dy \;\; P_X(x)$$

$$= \sum_{x=0}^{M-1} \int_{y \in A_x} f_{Y|X}(y \mid \hat{x}_A(y))\, P_X(\hat{x}_A(y))\, dy$$

$$= \int_{y=-\infty}^{\infty} f_{Y|X}(y \mid \hat{x}_A(y))\, P_X(\hat{x}_A(y))\, dy$$

$$= \int_{y=-\infty}^{\infty} P_{X|Y}(\hat{x}_A(y) \mid y)\, f_Y(y)\, dy$$

$$\leq \int_{y=-\infty}^{\infty} P_{X|Y}(\hat{x}_{MAP}(y) \mid y)\, f_Y(y)\, dy$$

$$= P(\hat{x}_{MAP}(Y) = X).$$

Application.

Abstraction of Digital Communication System

— Binary MAP Detection

$P_X(b) = p_1, \quad P_X(-b) = p_0$

$$Y = X + Z, \quad Z \sim N(\mu, \sigma^2)$$

X and Z are independent.

$$F_{Y|X=b}(y) = P(Y \leq y \mid X = b)$$

$$= \frac{P(X + Z \leq y \cap X = b)}{P(X = b)}$$

$$= \frac{P(Z \leq y - b)\,\cancel{P(X = x)}}{\cancel{P(X = b)}}$$

$$= F_Z(y - b)$$

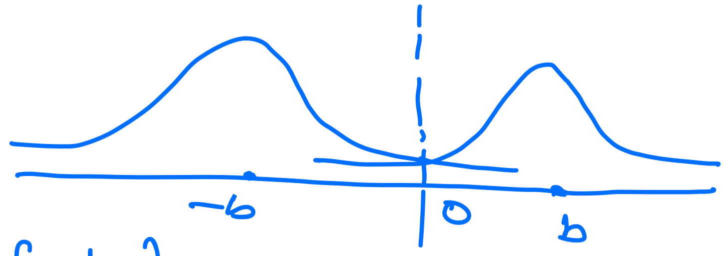
$$\Rightarrow f_{Y|X=b}(y) = f_Z(y - b)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-b)^2/2\sigma^2} =: f_{Y|X}(y|x).$$

$Y \mid X = b \sim N(b, \sigma^2)$.

similarly, $f_{Y|X}(y|-b) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(y+b)^2/2\sigma^2}$.

$Y|X=-b \sim \mathcal{N}(-b, \sigma^2)$.



MAP rule:

$$\hat{x}(y) = \arg\max_{x} \; p_{X|Y}(x|y)$$

$$= \arg\max_{x} \; \frac{f_{Y|X}(y|x)\, p_X(x)}{f_Y(y)}$$

$$= \arg\max_{x} \; f_{Y|X}(y|x)\, p_X(x)$$

$$= \begin{cases} b & \text{if } \; \Lambda(y) := \dfrac{f_{Y|X}(y|b)}{f_{Y|X}(y|-b)} \geq \dfrac{p_0}{p_1} \\[2ex] -b & \text{if } \; \Lambda(y) < \dfrac{p_0}{p_1} \end{cases},$$

$$e^{\frac{(y+b)^2-(y-b)^2}{2\sigma^2}} \;\underset{\hat{x}(y)=-b}{\overset{\hat{x}(y)=b}{\gtrless}}\; \frac{p_0}{p_1} =: \eta$$

$$\Rightarrow \; e^{2yb/\sigma^2} \;\underset{\hat{x}(y)=-b}{\overset{\hat{x}(y)=b}{\gtrless}}\; \eta$$

$$\Rightarrow \quad y \underset{\hat{x}(y)=-b}{\overset{\hat{x}(y)=b}{\gtrless}} \frac{\sigma^2}{2b} \log \eta$$

If $b=1$ and $p_0 = p_1$ this recovers the example on signal detection we have seen in an earlier lecture.

When $p_0 = p_1$ the decision rule is called Maximum Likelihood (ML) test. The ML test is often used when p_0 and p_1 are unknown.

Assuming $\eta = 1$ we find the overall probability of error.

$$P(\hat{x}(y) \neq X)$$

$$= P(\hat{x}(y) = -b \mid X = b)\, p_1 + P(\hat{x}(y) = b \mid X = -b)\, p_0$$

$$P(\text{Error} \mid X = b) = P(\hat{x}(y) = -b \mid X = b)$$

$$= P(Y < 0 \mid X = b)$$

$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-b)^2/2\sigma^2}\, dy$$

$$= \int_{-\infty}^{-b/\sigma} \frac{1}{\sqrt{2\pi}} e^{-y^2} dy$$

$$= P\left(N < \frac{-b}{\sigma}\right)$$

$$= \Phi(-b/\sigma).$$

$$= 1 - \Phi(b/\sigma).$$

$$P(\text{Error} \mid x = -b)$$

$$= P(\hat{x}(y) = b \mid x = -b)$$

$$= P(y > 0 \mid x = -b)$$

$$= \int_{0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y+b)^2/2\sigma^2} dy$$

$$= \int_{b/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= 1 - \Phi(b/\sigma).$$

$$P(\hat{x}(y) \neq x) = 1 - \Phi(b/\sigma).$$