Lecture 4

Question

- You catch a fish. Tell which one is it?
- Assumption: You cannot see the fish.

Decide only based on Prior: $P(\omega_1)$ or $P(\omega_2)$

$$P(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{\left(-\frac{1}{2}(x-\mu)^2/2\sigma^2\right)}$$

$x = \mu$

$$P(x = \mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$$

Cov. matrix Σ , $\mathrm{cov}(x) = \frac{1}{N-1}(X-\mu)(X-\mu)^T$

$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

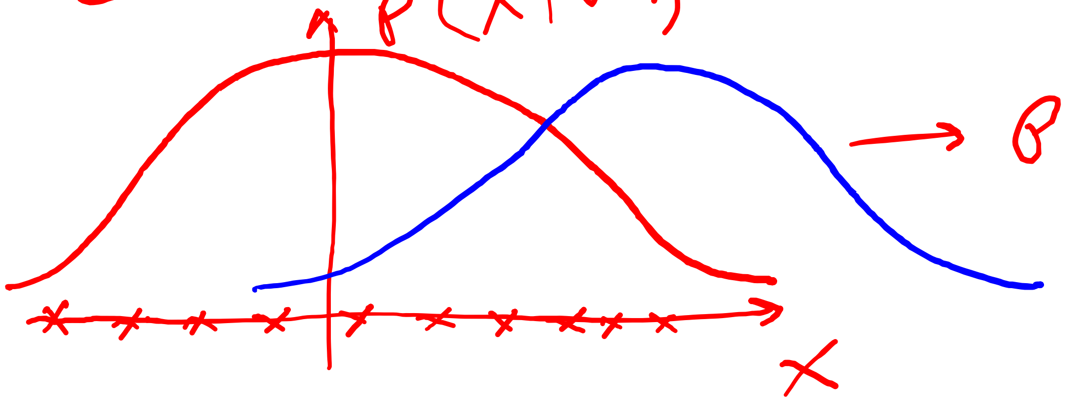
Question

- You catch a fish. You can see it, may be eat too. Tell which one is it?

M.V.G → $\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{\left\{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right\}}$ → Σ → P.D

X → weight of the fish.

B.D.T. $p(X|\omega_1) \sim N(\mu_1, \sigma_1^2)$



→ $p(X|\omega_2) \sim N(\mu_2, \sigma_2^2)$

$x \in \mathbb{R}$

$P(X|\omega_i)$ → class conditional

→ likelihood

~~For~~ $P(\omega_1|x)$, $P(\omega_2|x)$

Given: $P(\omega_1)$, $P(x|\omega_1)$, $P(\omega_1|x) = ?$

$P(\omega_2)$, $P(x|\omega_2)$, $P(\omega_2|x) = ?$

$P(\omega_1|x) \rightarrow$ Posterior Probability

$$P(\omega_1|x) = \frac{P(x|\omega_1)\,P(\omega_1)}{P(x)}$$

$$P(\omega_2|x) = \frac{P(x|\omega_2)\,P(\omega_2)}{P(x)}$$

Check:

$P(x|\omega_1)\,P(\omega_1) > P(x|\omega_2)\,P(\omega_2)$

Decide $\rightarrow \omega_1$

else $\rightarrow \omega_2$

- Decision rule with only the prior information
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$ otherwise decide ω_2

- Use of the class –conditional information

- $P(x \mid \omega_1)$ and $P(x \mid \omega_2)$ describe the difference in lightness between populations of sea and salmon

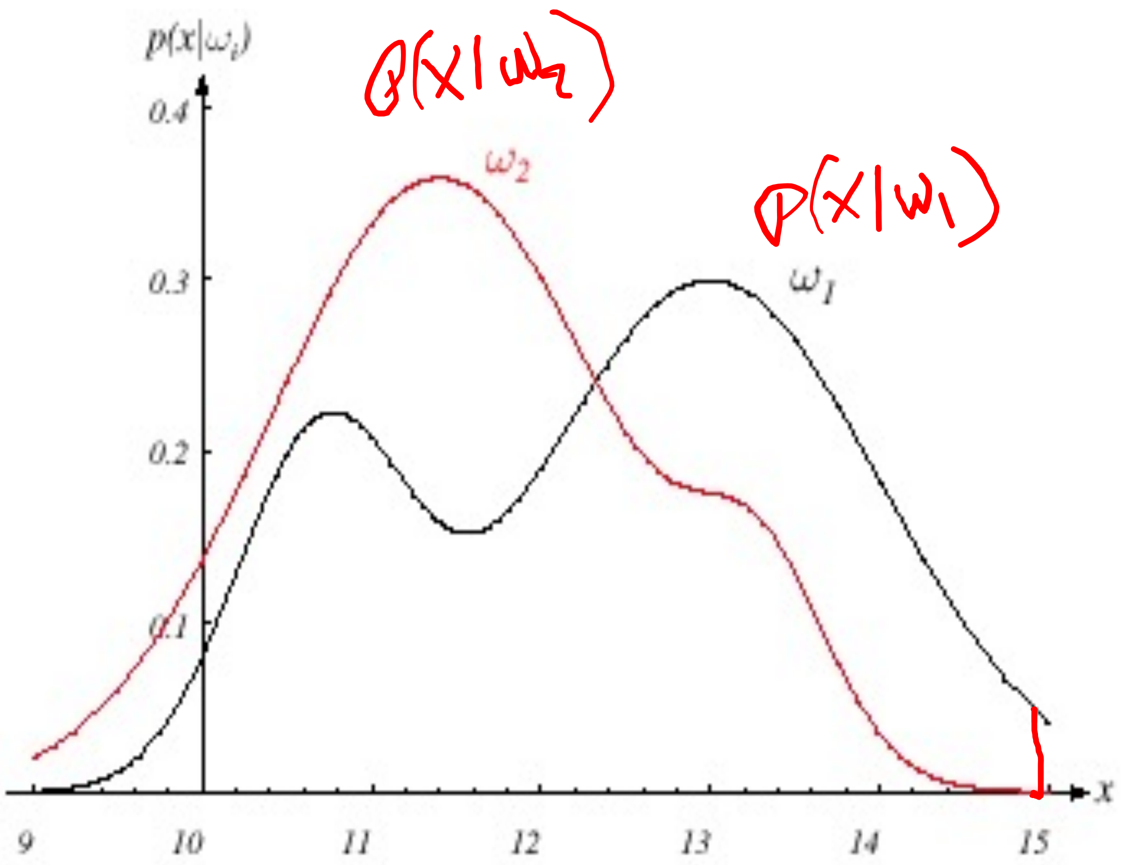


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- Posterior, likelihood, evidence

  - $P(\omega_j \mid x) = P(x \mid \omega_j) \,.\, P(\omega_j) / P(x)$

  - Where in case of two categories

$$P(x) = \sum_{j=1}^{j=2} P(x \mid \omega_j) P(\omega_j)$$

  - Posterior = (Likelihood. Prior) / Evidence

We have $P(\omega_i|x)$

$x = 9$, weight.

$P(\omega_2|x) > P(\omega_1|x)$

Decide : ω_2

G.7 → Seabass.

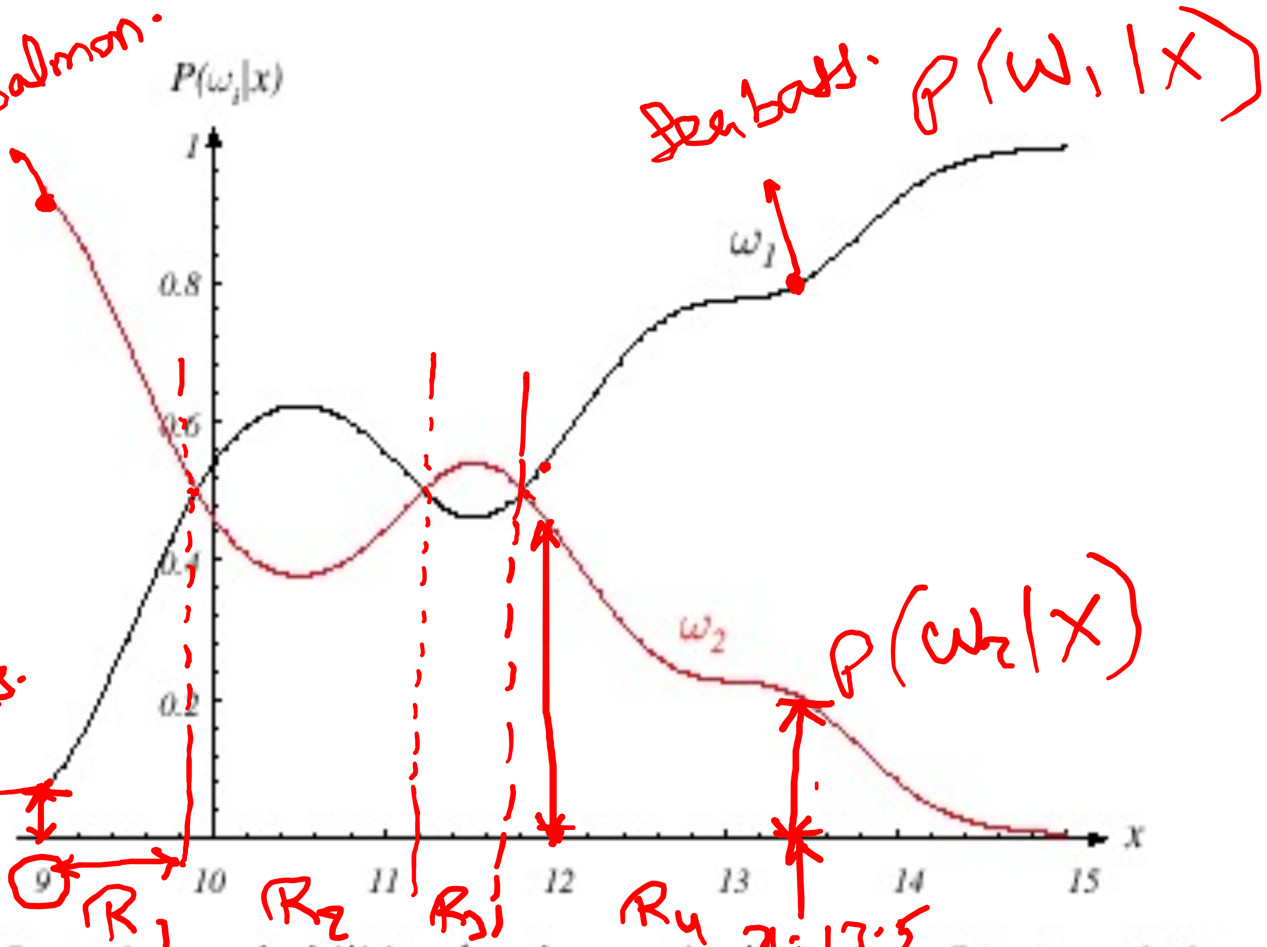


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

$P(\omega_1|x) = P(\omega_2|x)$ this x would give us decision boundary

$P(\text{error}|x) = ?$

G.7 → Salmon

- Decision given the posterior probabilities

X is an observation for which:

if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ $\Rightarrow$ True state of nature = ω_1

if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ $\Rightarrow$ True state of nature = ω_2

Therefore:

whenever we observe a particular x, the probability of error is :

$P(error \mid x) = P(\omega_1 \mid x)$ *if we decide* ω_2

$P(error \mid x) = P(\omega_2 \mid x)$ *if we decide* ω_1

- Minimizing the probability of error

- Decide ω_1 *if* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$; otherwise decide ω_2

Therefore:

$$P(error \mid x) = \min\,[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

(Bayes decision)

$$P(error) = \int_x \underbrace{P(error \mid x)} \, p(x)\, dx$$

$$= \int_{x \in R_1} \underbrace{P(\omega_1 \mid x)\, p(x)\, dx}_{p(x \mid \omega_1) P(\omega_1)\, dx}$$

$$+ \int_{x \in R_2} P(\omega_2 \mid x)\, p(x)\, dx$$

$$+ \int_{x \in R_3} P(\omega_1 \mid x)\, p(x)\, dx$$

$$+ \int_{x \in R_4} P(\omega_2 \mid x)\, p(x)\, dx$$

Show that for arbitrary densities, we can replace Eq. 7 by $P(error|x) = 2P(\omega_1|x)P(\omega_2|x)$ in the integral and get an upper bound on the full error.

$P(error|x) \rightarrow$ max value of this is bounded

$e = 2P(\omega_1|x)P(\omega_2|x)$

$P(\omega_1|x) = a$

$a < 1/2$

$$\underbrace{P(error|x)} < \underbrace{2P(\omega_1|x)P(\omega_2|x)}$$

$\min(P(\omega_1|x), P(\omega_2|x)) <$,, max. $P(error|x)$

$\min(a, 1-a) < 2a(1-a)$

$a < 2a(1-a) = 2a - 2a^2$

$a < 2a - 2a^2$ $\quad a < 1/2$

Bayesian Decision Theory – Continuous Features

$x \in R$

- Generalization of the preceding ideas
  - Use of more than one feature → $x \in R^d$
  - Use more than two states of nature → $\omega_1, \omega_2, \omega_3, \ldots \omega_c$
  - Allowing actions and not only decide on the state of nature
  - Introduce a loss of function which is more general than the probability of error

- Allowing actions other than classification primarily allows the possibility of rejection

- Refusing to make a decision in close or bad cases!

- The loss function states how costly each action taken is

Let $\{\omega_1, \omega_2, \ldots, \omega_c\}$ be the set of c states of nature (or "categories")

Let $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ be the set of possible actions

Let $\lambda(\alpha_i \mid \omega_j)$ be the loss incurred for taking

action α_i when the state of nature is ω_j

Overall risk

R = Sum of all $\underbrace{R(\alpha_i \mid x)}_{\textbf{Conditional risk}}$ *for i = 1,…,a*

$R(\alpha_i \mid x)$

$P(\omega_i \mid x)$

Minimizing R ⟺ Minimizing $R(\alpha_i \mid x)$ *for i = 1,…, a*

$$R(\alpha_i \mid x) = \sum_{j=1}^{j=c} \underbrace{\lambda(\alpha_i \mid \omega_j)}_{\text{loss}} \underbrace{P(\omega_j \mid x)}$$

$= \sum_{j=1}^{2} \lambda(\alpha_i \mid \omega_j) \, P(\omega_j \mid x)$

$R(\alpha_i \mid x)$

for i = 1,…,a

Select the action α_i for which $R(\alpha_i \mid x)$ is minimum

R is minimum and R in this case is called the
Bayes risk = best performance that can be achieved!

- Two-category classification

α_1 : deciding ω_1

α_2 : deciding ω_2

$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$

$\lambda_{11} = 0 = \lambda_{22}$

$\lambda_{12}, \lambda_{11} > 0$

loss incurred for deciding ω_i when the true state of nature is ω_j

Conditional risk:

Risk for action α_1

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$

$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

If $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$: Decide ω_1

else: Decide ω_2

Our rule is the following:

$$\text{if } R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$$

action α_1: "decide ω_1" is taken

This results in the equivalent rule :

decide ω_1 if:

$$\lambda_{11} P(\omega_1 | x) + \lambda_{12} P(\omega_2 | x) <$$
$$\lambda_{21} P(\omega_1 | x) + \lambda_{22} P(\omega_2 | x)$$

$$(\lambda_{21} - \lambda_{11}) P(x \mid \omega_1) P(\omega_1) > (\lambda_{12} - \lambda_{22}) P(x \mid \omega_2) P(\omega_2)$$

and decide ω_2 otherwise

Likelihood ratio:

The preceding rule is equivalent to the following rule:

$$if \; \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > \underbrace{\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}} \; = \frac{1}{2}$$

Ratio of likelihood.

Then take action α_1 (decide ω_1)

Otherwise take action α_2 (decide ω_2)

Optimal decision property

"If the likelihood ratio exceeds a threshold value independent of the input pattern x, we can take optimal actions"

Exercise

Select the optimal decision where:

⍰ = $\{\omega_1, \omega_2\}$

$P(x \mid \omega_1)$ N(2, 0.5) (Normal distribution)

$P(x \mid \omega_2)$ N(1.5, 0.2)

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

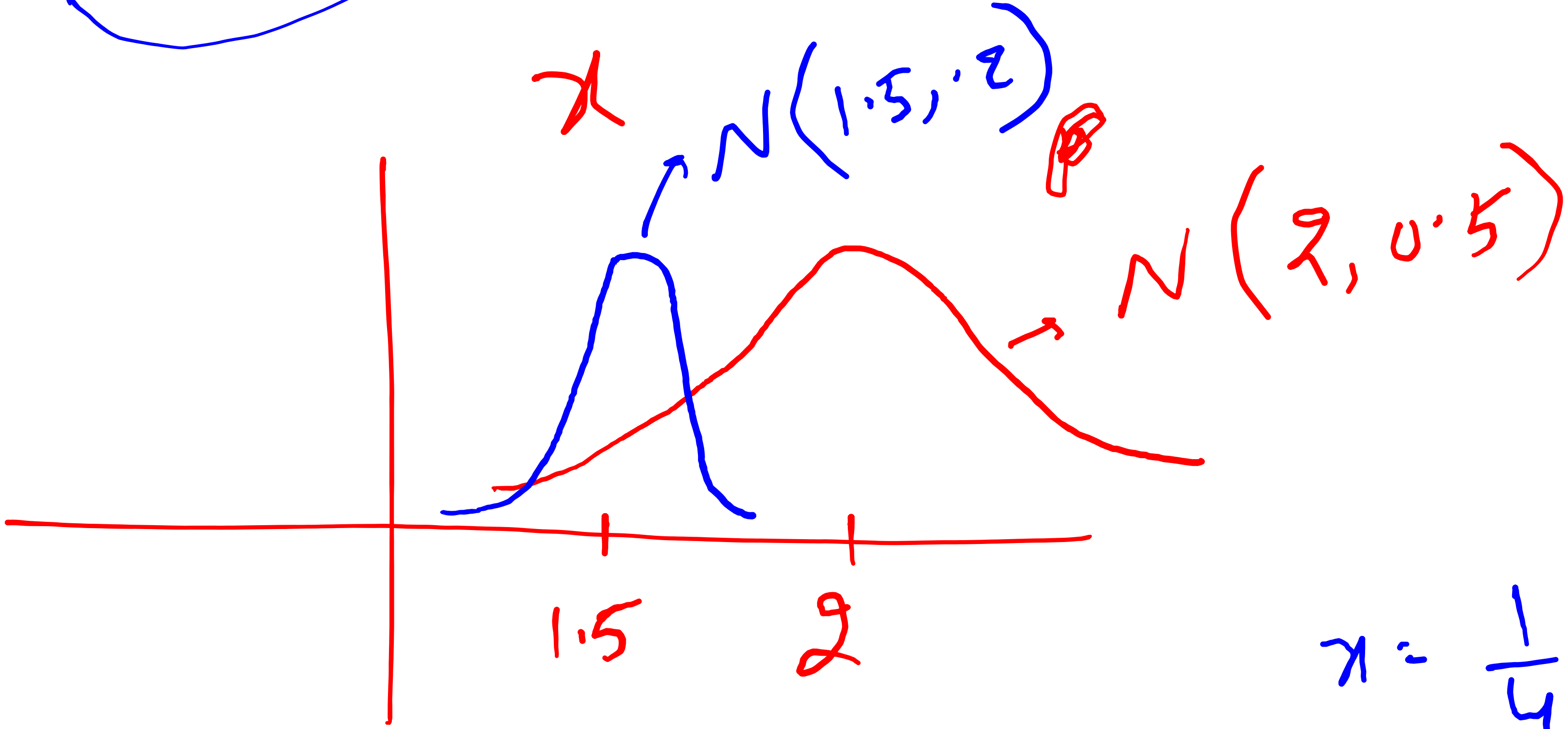
$$\lambda = \cancel{\begin{bmatrix} \overset{\lambda_{11}}{1} & 2 \\ 3 & \underset{\lambda_{22}}{4} \end{bmatrix}}$$

$$\frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} = \frac{\dfrac{1}{\sqrt{2\pi \cdot 1/2}} e\left\{-\dfrac{1}{2}\dfrac{(x-2)^2}{1/2}\right\}}{\dfrac{1}{\sqrt{2\pi \cdot 1/5}} e\left\{-\dfrac{1}{2}\dfrac{(x-3/2)^2}{1/5}\right\}}$$

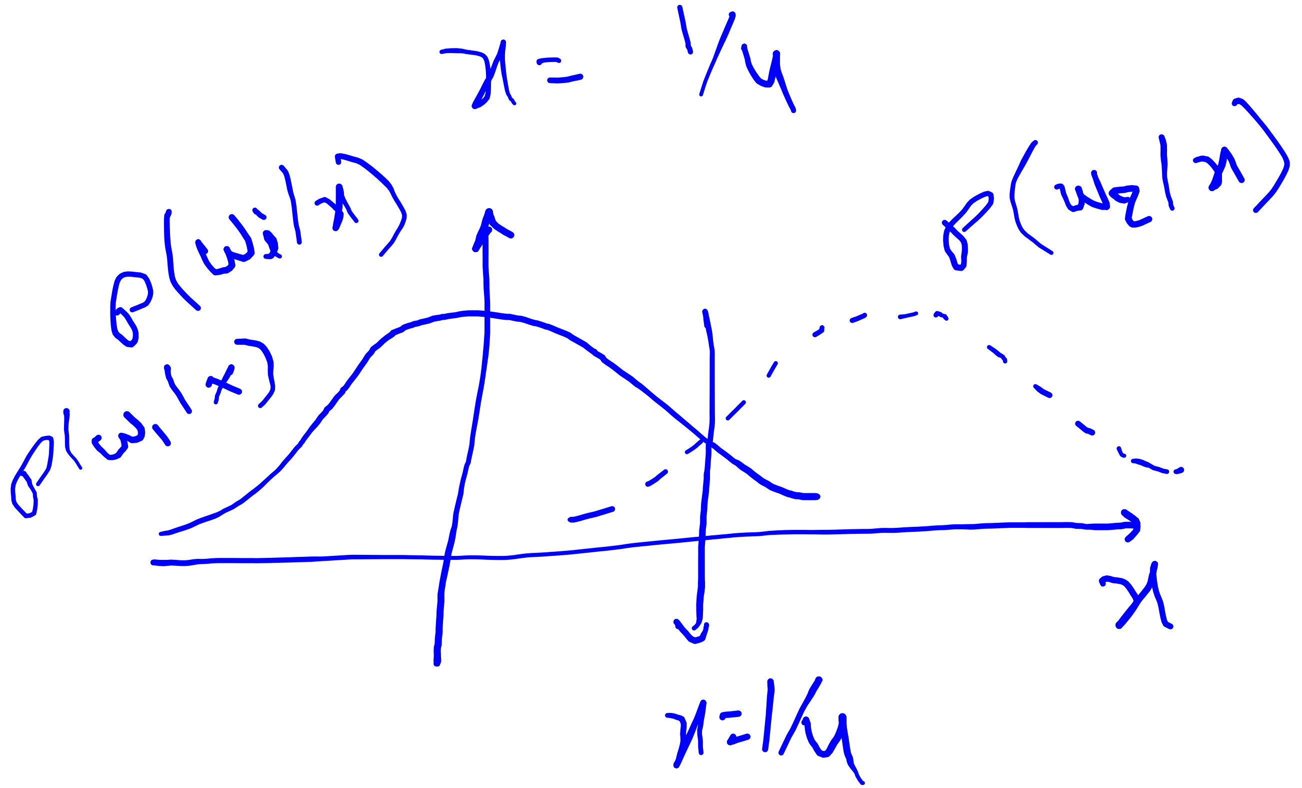
$$= \sqrt{\frac{2}{5}}\, e\left\{+\frac{5}{2}\left(x-\frac{3}{2}\right)^2 - (x-2)^2\right\} > 1/2$$

$$\lambda = \cancel{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} \qquad \lambda = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$$

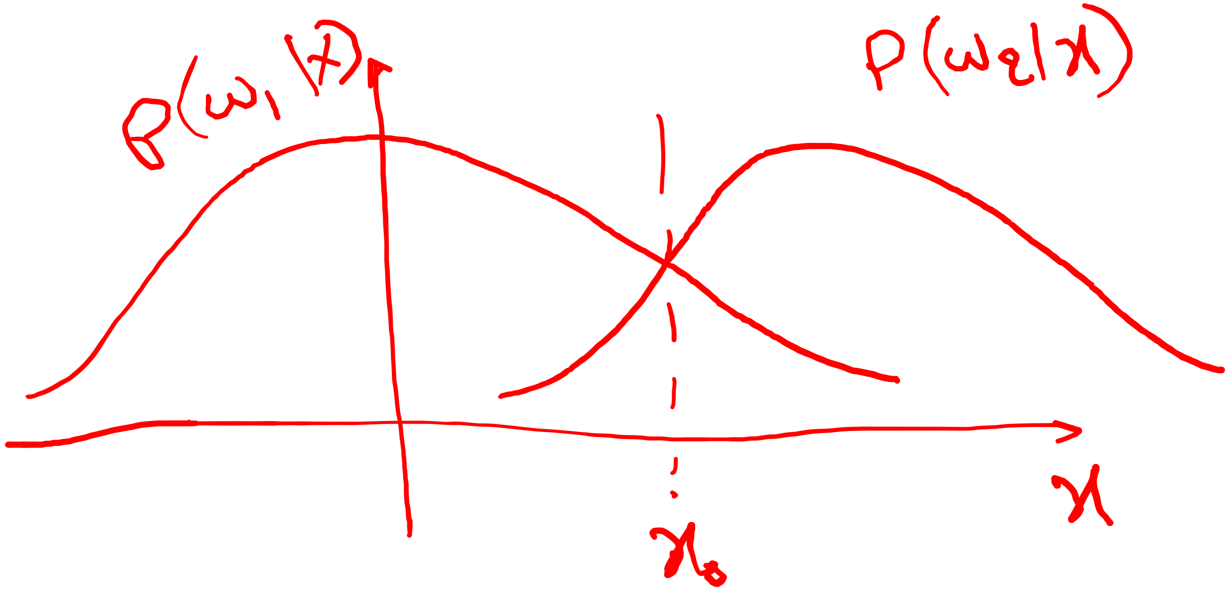
$$\underbrace{\ln\frac{\sqrt{2}}{5} + \frac{5}{2}\left\{\left(x-\frac{3}{2}\right)^2 - (x-2)^2\right\}} > \ln 1/2$$



$x = \frac{1}{4}$



$x = 1/4$
$p(\omega_1|x)$
$p(\omega_1|x)$
$p(\omega_2|x)$
x
$x = 1/4$



D.B. $x = x_0$

$$P(\text{error}) = ?$$

$$= \int_{-\infty}^{x_0} P(\omega_2|x)\,P(x)\,dx + \int_{x_0}^{\infty} P(\omega_1|x)\,P(x)\,dx$$

Decision boundary.

$$P(\omega_1|x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\left\{-\frac{1}{2}(x-1)^2/\sigma^2\right\}} \qquad \sigma = 1$$

$$P(\omega_2|x) = \frac{1}{\sqrt{2\pi\sigma^2}}\left\{ e^{-\frac{1}{2}(x-2)^2/\sigma^2}\right\}$$

$$P(\text{error}|x) = \min\left[P(\omega_1|x),\ P(\omega_2|x)\right]$$

Chapter 2 (Part 2): Bayesian Decision Theory (Sections 2.3-2.5)

Minimum-Error-Rate Classification

Classifiers, Discriminant Functions and Decision Surfaces

The Normal Density

Minimum-Error-Rate Classification

$\int p(\omega|x)\,p(x)\,dx$

$\int p(x|\omega)\,p(\omega)\,\cancel{p(x)}\,\cancel{p(x)}\,dx$

- Actions are decisions on classes

  If action α_i is taken and the true state of nature is ω_j then:

  the decision is correct if $i = j$ and in error if $i \neq j$

- Seek a decision rule that minimizes the *probability of error* which is the *error rate*

D.R. $\min. R(\alpha_i|x)$

$\max. P(\omega_i|x)$

- Introduction of the zero-one loss function:

loss or reject. λ_r

$$\lambda(\alpha_i,\omega_j)=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases}\qquad i,j=1,\dots,c$$

$\lambda(\alpha_{c+1}) \leftarrow$ ~~decide no class~~ λ_r

Therefore, the conditional risk is:

$$\underbrace{R(\alpha_i\,|\,x)}_{\min}=\sum_{j=1}^{j=c}\lambda(\alpha_i\,|\,\omega_j)P(\omega_j\,|\,x)$$

$$=\sum_{j\neq 1}P(\omega_j\,|\,x)=1-\underbrace{P(\omega_i\,|\,x)}_{\max.}$$

$\lambda(\alpha_i|\omega_1)P(\omega_1|x)$
$+\lambda(\alpha_i|\omega_2)P(\omega_2|x)$
$+\cancel{\lambda(\alpha_i|\omega_i)P(\omega_i|x)}$
$+\lambda(\alpha_i|\omega_c)P(\omega_c|x)$

$\sum_{j\neq i}P(\omega_j|x)$

"The risk corresponding to this loss function is the average probability error"

$P(x \mid \omega_1)$ ~~P(x~~

↓ class conditional pdf

$P(x = x_i \mid \omega_1)$

ω_1

$x_i = \{1, 2, 3, 4\}$

$P(x = 1 \mid \omega_1)$

- Minimize the risk requires maximize $P(\omega_i \mid x)$ (since $R(\alpha_i \mid x) = 1 - P(\omega_i \mid x)$)

ω_2

$x_i = \{2.5, 6, 10\}$

- For Minimum error rate
  - Decide ω_i if $P(\omega_i \mid x) > P(\omega_j \mid x) \ \forall j \neq i$