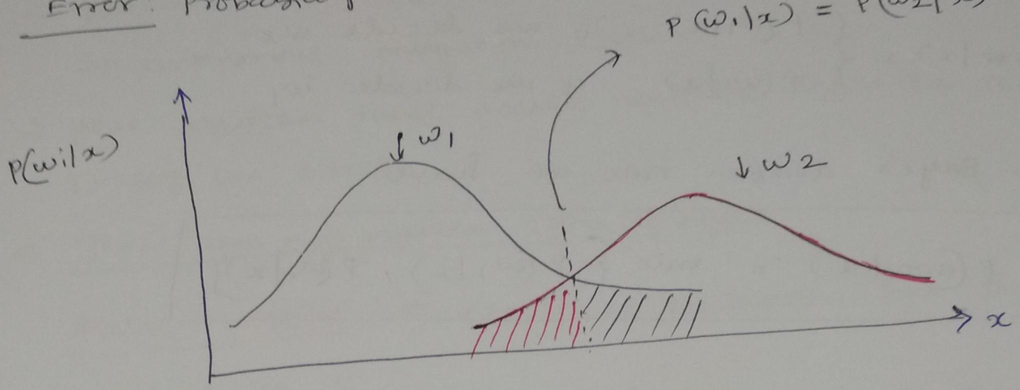
30/1/2019

Error: Probability of error



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

There is a finite probability of error

If we decide ~~for~~ in favour of $\omega_1 \Rightarrow P(\omega_2|x)$

If we decide ~~for~~ in favour of $\omega_2 \Rightarrow P(\omega_1|x)$

Given The situation like this, the total error

$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error}, x)\, dx$$

Joint probability

$$= \int_{-\infty}^{\infty} P(\text{error}|x) \cdot P(x)\, dx$$

$$= \boxed{\min \{ P(\omega_1|x),\ P(\omega_2|x) \}}$$

II

Whenever, we observe a particular x, the probability of error is

$$P(\text{error}|x) = \begin{cases} P(\omega_1|x) & \text{if we decide } \omega_2 \\ P(\omega_2|x) & \text{if we decide } \omega_1 \end{cases}$$

under Baye's decision rule, we have

$$\boxed{P(\text{error}|x) = \min\{P(\omega_1|x), P(\omega_2|x)\}}$$

Is Baye's Decision rule optimal?

* For every x, we ensure that $P(\text{error}|x)$ is as small as possible.

* The "average probability of error" over all possible x, must also be as small as possible.

In general, for 'C' classes

$$\boxed{P(\omega_i|x) > P(\omega_j|x) \quad \forall i \neq j \Rightarrow \omega_i}$$

1. The Baye's rule minimizes the expected error rate.
2. Minimizing the expected error rate is a pretty reasona[ble] goal.
3. However, it is not always the best thing to do.

Example:

→ You are designing a pedestrain detection algorithm for an autonomous navigation system.

→ your algorithm must decide whether there is a pedestrain crossing the street.

⇒ There are two possible types of error:

False positive: There is no pedestrain, but the Algo thinks there is

Miss: There is a pedestrain, but the Algo thinks there is not.

⇒ Should we give equal weight to these 2 types of error?

sol: To deal with this problem, instead of minimizing error rate, we minimize something called the risk.

1. First, we define the loss matrix L, which quantifies the cost of making each type of error

2. Element λ_{ij} of the loss matrix specifies the cost of deciding class j when in fact the input is of class i.

(or)

Loss incurred for taking action α_i when true state of nature is ω_j.

3. Typically, we set $\lambda_{ii} = 0$ for all i.

... is diagonal

Thus a typical loss matrix for $M=2$; would have the form

$$L = \begin{bmatrix} 0 & \lambda_{12} \\ \lambda_{21} & 0 \end{bmatrix} - \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$$

Baye's Risk : [Baye's Decision rule – The General case:]

[~~Baye's minimum Risk classifier:~~]

There can be a generalization of this Baye's Theorem

1. use more than two states of nature ⇒ classes
2. use more than one feature for observation ⇒ Feature vector
3. Allow other actions other than merely deciding states of nature
4. Introduce a loss function more general than the probability of error.

i) C – state of nature

$\Omega = \{ \omega_1, \omega_2, \ldots \omega_c \}$

ii) X ⇒ d-dimensional feature vector

iii) a – Actions

$A = \{ \alpha_1, \alpha_2, \ldots \alpha_a \}$

loss function : lamda (alpha$_i$ | omega$_j$)

$\lambda(\alpha_i | \omega_j) \Rightarrow$ loss incurred for taking action α_i when true state of nature is ω_j. It can be written as λ_{ij}.

now, given these, how to ~~decide~~ take the decision rule in these generalized Baye's Theory. ?

given:
1. C - number of classes
2. x_d
3. Take an action α_i
4. loss function $\lambda(\alpha_i | \omega_j) = \lambda_{ij}$

If the probability that the true states of nature is ω_j given the feature vector $x \Rightarrow P(\omega_j / x)$

Then the average risk / loss can be computed as

$$\boxed{R(\alpha_i / x) = \sum_{j=1}^{C} \lambda(\alpha_i | \omega_j) \cdot P(\omega_j / x)}$$

Expected loss is also called Risk function | conditional Risk.

now, I have to choose that action α_i for which the risk is minimum. also called as Baye's Risk, and is the best performance that can be achieved.

Special case:

let us see for two category cases: suppose, I have two classes ω_1 and ω_2

$$\Omega = \{\omega_1, \omega_2\}$$

$$A = \{\alpha_1, \alpha_2\} \begin{cases} \alpha_1 \Rightarrow \text{decide to } \omega_1 \\ \alpha_2 = \text{decide to } \omega_2 \end{cases}$$

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

$$\boxed{R(\alpha_i / x) = \sum_{j=1}^{C} \lambda(\alpha_i/\omega_j) \cdot P(\omega_j/x)}$$

$$\boxed{\begin{aligned} R(\alpha_1 | x) &= \lambda_{11} \cdot P(\omega_1/x) + \lambda_{12} \cdot P(\omega_2/x) \\ R(\alpha_2 | x) &= \lambda_{21} \cdot P(\omega_1|x) + \lambda_{22} \cdot P(\omega_2|x) \end{aligned}}$$

Now, I have to choose that action α_i for which the risk is minimum.

If $R(\alpha_1/x) < R(\alpha_2|x) \Rightarrow$ take action α_1 ~~cad~~ decide cla[ss] ω_1

31/1/2019

Minimum-Error-Rate classification:

If I am taking an action α_i, I am taking a decision true state of nature is ω_i

$\alpha_i \rightarrow$ true state of nature is ω_i.

If I define the loss fn.

$$\lambda(\alpha_i/\omega_j) = \begin{cases} 0 & i=j \\ 1 & i \neq j \end{cases} \quad i,j = 1,2,\dots c$$

By this definition of loss fn, what will be the expected loss (or) Risk.

$$R(\alpha_i/x) = \sum_{j=1}^{c} \lambda(\alpha_i/\omega_j) \cdot P(\omega_j/x)$$

$$= \underbrace{\lambda(\alpha_1/\omega_1)} \cdot P(\omega_1/x) + \underbrace{\lambda(\alpha_1/\omega_2)}_{1} \cdot P(\omega_2/x) + \dots \underbrace{\lambda(\alpha_1/\omega_c)}_{1} \cdot P(\omega_c/$$

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

$$\Downarrow$$

This is nothing but

$$\boxed{R(\alpha_i/x) = 1 - P(\omega_i/x)}$$

$\Uparrow$

$1 - P(\omega_i/x)$

error rate

$\Rightarrow$ The probability $P(\omega_i/x)$ that action α_i (decide ω_i) is correct

$\boxed{1 - P(\omega_i/x)}$ is wrong

So, If I want to minimize this risk for $\boxed{1 - P(\omega_j)}$ has to be minimum. If I want to minimise $\boxed{1 - P(\omega}$

then $\boxed{P(\omega_j / x)}$ has to be maximum.

Now, I ~~come~~ come back similar decision in the generalized case, for which ever j is maximum, I choose that. $\rightarrow$ $P(\omega_j / x)$

Baye's classifier is

$$\boxed{\text{Decide } \omega_i \text{ If } P(\omega_i / x) > P(\omega_j / x) \text{ for all } i \neq j}$$

$\Downarrow$

Thus, to minimize the average probability of error, we should select the i that maximizes the posterior probability $P(\omega_i / x)$. In other words for minimum error rate

Example

Baye's

Example 1:

Baye's Decision Rule – The general case:

Action / Class	α_1 Recipe A	α_2 Recipe B	α_3 No Recipe
ω_1 = cancer	5	50	10,000
ω_2 = No cancer	60	3	0

For a particular x:

$P(\omega_1/x) = 0.01$

$P(\omega_2/x) = 0.99$

$$R(\alpha_1|x) = \sum_{j=1}^{c=2} \lambda(\alpha_i/\omega_j) \cdot P(\omega_j/x)$$

$$= \lambda(\alpha_1/\omega_1) \cdot P(\omega_1|x) + \lambda(\alpha_1/\omega_2) \cdot P(\omega_2|x)$$

$$= (5 \times 0.01) + (60 \times 0.99)$$

$$= \boxed{59.45}$$

$$R(\alpha_2|x) = \lambda(\alpha_2/\omega_1) \cdot P(\omega_1|x) + \lambda(\alpha_2/\omega_2) \cdot P(\omega_2|x)$$

$$= (50 \times 0.01) + (3 \times 0.99)$$

$$= 0.5 + 2.97 = 3.47$$

$$= \boxed{3.47}$$

$$\lambda(\alpha_3|\omega_1) \cdot P(\omega_1|x) + \lambda(\alpha_3|\omega_2) \cdot P(\omega_2|x)$$

$$R(\alpha_3|x) = 10{,}000 \times 0.01 + 0 \times 0.99$$

$$= \boxed{100}$$

Among all, $R(\alpha_2|x)$ is minimum (Risk is minimum)

then perform action α_2 and decide class ω_2.

take 'Recipe B' and decide 'no cancer'.

Example 2:

SPAM filtering: => turn next page (7)

i) There are two actions

$A = \{\alpha_1, \alpha_2\}$

$\alpha_1 \Rightarrow$ stands for keep the mail

$\alpha_2 \Rightarrow$ stands for delete as SPAM.

ii) There are two classes

$\Omega = \{\omega_1, \omega_2\}$

$\omega_1 \Rightarrow$ normal mail

$\omega_2 \Rightarrow$ SPAM i.e junk mail

(7)
(6)

iii) $P(\omega_1) = 0.4$ | $P(\omega_2) = 0.6$

iv) Loss fn [matrix]

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 3 & 0 \end{bmatrix}$$

λ_{21} is costly (i.e 3); that is deleting important

(α_2/ω_1) mail as SPAM is more costly than keeping as SPAM mail

α_2 - deleting mail as SPAM

ω_1 - actually it belongs to class $\omega_1 \Rightarrow$ normal mail.

v) we get an E-mail message with the feature vector X and $P(X/\omega_1) = 0.35$, $P(X/\omega_2) = 0.65$

How does the Baye's minimum risk classifier act?

Solution:

$$P(\omega_1 | X) = \frac{P(X|\omega_1) \cdot P(\omega_1)}{P(X)}$$

$$= \frac{0.35 \times 0.4}{(0.35 \times 0.4) + (0.65 \times 0.6)} = 0.264.$$

$P(\omega_2/x) = 1 - 0.264$

$= 0.736$

$R(\alpha_1/x) = \underline{\lambda(\alpha_1/\omega_1)} \cdot P(\omega_1/x) + \underline{\lambda(\alpha_1/\omega_2)} \cdot P(\omega_2/x)$

$= (0 \times 0.264) + (1 \times 0.736)$

$= 0.736$

$R(\alpha_2/x) = \underline{\lambda(\alpha_2/\omega_1)} \cdot P(\omega_1/x) + \lambda(\alpha_2/\omega_2) \cdot P(\omega_2/x)$

$= (3 \times 0.264) + (0 \times 0.736)$

$= 0.792$

since $R(\alpha_1/x) < R(\alpha_2/x)$, we decide to take action α_1 and decide class ω_1.

$\Downarrow$ Keep the mail

$\Downarrow$ normal mail.

$\boxed{\text{(i.e) Don't delete the mail.}}$

Given:

Class \ Action	α_1 Keep the mail	α_2 Delete as SPAM
ω_1 = normal mail	0	3
ω_2 = SPAM mail	1	0

$\lambda(\alpha_i | \omega_j)$

ii) $P(\omega_1) = 0.4$ | $P(\omega_2) = 0.6$

iii) $P(x|\omega_1) = 0.35$, $P(x|\omega_2) = 0.65$

How does the Baye's minimum risk classifier act?

①

A

1. Bayes Minimum Error Classifier
(Mimimizing the probablity of error)

Clip slide

Bayes' Theorem: Example

- Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively. The first box (B_1) has 15 defective bulbs and the second has 5 defective bulbs. Suppose a box is selected at random and one bulb is picked out.
  a) What is the probability it is defective?
  b) Suppose the bulb we tested was defective. What is the probability it came from box 1.

$x_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$x_2 = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$

$$\boxed{a_k(x) = x_k^T \Sigma_{pooled}^{-1} x - \frac{1}{2} x_k^T \Sigma_p^{-1} x_k + \ln \frac{N_k}{N}}$$

$$g_1(x) = [2 \; 2] \frac{1}{4} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \frac{1}{8} [2 \; 2] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \ln \left\{ \frac{1}{2} \right\}$$

$$g_2(x) = \frac{1}{4}$$

$$\boxed{\Sigma_{pooled} = P(\omega_1) \Sigma_1 + P(\omega_2) \Sigma_2}$$

- Solution: Part (a)

Note that box B_1 has 85 good and 15 defective bulbs. Similarly box B_2 has 195 good and 5 defective bulbs.

B_1 and B_2 form the portions of the box.
Let D = "Defective bulb is picked out".

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$

$$P(D \mid B_1) = \frac{15}{100} = 0.15, \quad P(D \mid B_2) = \frac{5}{200} = 0.025.$$

Since the box is selected at random, they are equally likely

$$P(B_1) = P(B_2) = \frac{1}{2}$$

The probability of event D is:

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$

$$P(D) = (0.15)\left(\frac{1}{2}\right) + (0.025)\left(\frac{1}{2}\right) = 0.0875$$

Thus, there is about 9% probability that a bulb picked at random is defective

$$P(B_1 \mid D) = \frac{P(D/B_1) \cdot P(B_1)}{P(D)}$$

$$= \frac{0.15 \times 0.5}{0.0875} = \underline{0.857}$$

- Part (b): Suppose the bulb we tested was defective. What is the probability it came from box 1? $P(B_1 \mid D) = ?$

- Solution:

Notice that initially $P(B_1) = P(B_2) = 0.5$, then we picked out a box at random and tested a bulb that turned out to be defective. Can this information shed some light about the fact that we might have picked up box 1?

$$P(B_1 \mid D) = 0.857 > 0.5$$

Indeed it is more likely at this point that we must have chosen box 1 in favor of box 2. (Recall box1 has three times more defective bulbs compared to box2).

Note/ Calculation:

P(B1/D)= P(D/B1)* P(B1)/ P(D)
= 0.15*0.5/0.0875
= 0.8571

Another Example

Problem statement

- A new medical test is used to detect whether a patient has a certain cancer or not, whose test result is either + (*positive*) or - (*negative*)
- For patient with this cancer, the probability of returning *positive* test result is 0.98
- For patient without this cancer, the probability of returning *negative* test result is 0.97
- The probability for any person to have this cancer is 0.008

Question

If *positive* test result is returned for some person, does he/she have this kind of cancer or not?

$$P(\omega_i / +) = ? \Rightarrow \begin{array}{l} P(\omega_1 | +) = ? \\ P(\omega_2 | +) = ? \end{array}$$

If $P(\omega_1 | +) > P(\omega_2 | +) \Rightarrow \omega_1$

otherwise $\Rightarrow \omega_2$.

Another Example (Cont.)

ω_1 : cancer $\qquad \omega_2$: no cancer $\qquad x \in \{+, -\}$

$P(\omega_1) = 0.008 \qquad P(\omega_2) = 1 - P(\omega_1) = 0.992$

$P(+ \mid \omega_1) = 0.98 \qquad P(- \mid \omega_1) = 1 - P(+ \mid \omega_1) = 0.02$

$P(- \mid \omega_2) = 0.97 \qquad P(+ \mid \omega_2) = 1 - P(- \mid \omega_2) = 0.03$

$$P(\omega_1 \mid +) = \frac{P(\omega_1)P(+ \mid \omega_1)}{P(+)} = \frac{P(\omega_1)P(+ \mid \omega_1)}{P(\omega_1)P(+ \mid \omega_1) + P(\omega_2)P(+ \mid \omega_2)}$$

$$= \frac{0.008 \times 0.98}{0.008 \times 0.98 + 0.992 \times 0.03} = 0.2085$$

$$P(\omega_2 \mid +) = 1 - P(\omega_1 \mid +) = 0.7915$$

$\boxed{P(\omega_2 \mid +) > P(\omega_1 \mid +)}$

No cancer!

x

	+	−
ω_1 cancer	0.98	
no cancer ω_2		0.97

Given : $P(+/\omega_1) = 0.98 \Rightarrow P(-/\omega_1) = 1 - 0.98 = 0.02$

$P(-/\omega_2) = 0.97 \Rightarrow P(+/\omega_2) = 1 - 0.97 = 0.03$

$P(\omega_1) = 0.008 \Rightarrow P(\omega_2) = 1 - 0.008 = 0.992$

Is Bayes Decision Rule Optimal?

Bayes Decision Rule (In case of two classes)

If $P(\omega_1|x) > P(\omega_2|x)$, Decide ω_1; Otherwise ω_2

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

$$P(\text{error} \mid x) = \begin{cases} P(\omega_1 \mid x) & \text{if we decide } \omega_2 \\ P(\omega_2 \mid x) & \text{if we decide } \omega_1 \end{cases}$$

Under Bayes decision rule, we have

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

For every x, we ensure that $P(\text{error} \mid x)$ is as small as possible → The average probability of error over all possible x must be as small as possible

2. Bayes Minimum Risk Classifier

Bayes Decision Rule – The General Case (Cont.)

- By introducing a loss function more general than the probability of error

$$\lambda : \Omega \times A \rightarrow \mathbf{R} \quad \text{(loss function)}$$

$\lambda(\omega_j, \alpha_i)$: the loss incurred for taking action α_i when the state of nature is ω_j

For ease of reference, usually written as: $\lambda(\alpha_i \mid \omega_j)$

A simple loss function

Class \ Action	α_1 = "Recipe A"	α_2 = "Recipe B"	α_3 = "No Recipe"
ω_1 = "cancer"	5	50	10,000
ω_2 = "no cancer"	60	3	0

Action α_i decides w_i

$$\lambda(\alpha_i / w_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad i, j = 1, 2, 3 \ldots c$$

$$R(\alpha_i / x) = \sum_{j=1}^{c} \underbrace{\lambda(\alpha_i / w_j)} \cdot P(w_j / x)$$

$$= \sum_{i \neq j} P(w_j / x)$$

$$\boxed{1 - P(w_i / x)}$$

decide

$P(w_i / x) > P(w_j / x)$ for all $i \neq j$.

Bayes Decision Rule – The General Case (Cont.)

Suppose we have:

Class \ Action	$\alpha_1 =$ "Recipe A"	$\alpha_2 =$ "Recipe B"	$\alpha_3 =$ "No Recipe"
$\omega_1 =$ "cancer"	5	50	10,000
$\omega_2 =$ "no cancer"	60	3	0

For a particular $\mathbf{x}$:
$P(\omega_1 \mid \mathbf{x}) = 0.01$
$P(\omega_2 \mid \mathbf{x}) = 0.99$

$$R(\alpha_1 \mid \mathbf{x}) = \sum_{j=1}^{2} \lambda(\alpha_1 \mid \omega_j) \cdot P(\omega_j \mid \mathbf{x})$$
$$= \lambda(\alpha_1 \mid \omega_1) \cdot P(\omega_1 \mid \mathbf{x}) + \lambda(\alpha_1 \mid \omega_2) \cdot P(\omega_2 \mid \mathbf{x})$$
$$= 5 \times 0.01 + 60 \times 0.99 = 59.45$$

Similarly, we can get: $R(\alpha_2 \mid \mathbf{x}) = 3.47$ $\quad R(\alpha_3 \mid \mathbf{x}) = 100$

Another Example:

Example: SPAM filtering

- We have two actions: α_1 stands for **keep the mail** and α_2 stands for **delete as SPAM**. There are two classes ω_1 (normal mail) ω_2 (SPAM i.e. junk mail).
- $P(\omega_1) = 0.4, P(\omega_2) = 0.6$ and $\lambda_{11} = 0, \lambda_{21} = 3$, $\lambda_{12} = 1, \lambda_{22} = 0$. That is, deleting important mail as SPAM is more costly than keeping a SPAM mail.
- We get an e-mail message with the feature vector x and $p(\mathbf{x}|\omega_1) = 0.35, p(\mathbf{x}|\omega_2) = 0.65$. How does the Bayes minimum risk classifier act?
- $P(\omega_1|\mathbf{x}) = \frac{0.35 \cdot 0.4}{0.35 \cdot 0.4 + 0.65 \cdot 0.4} = 0.264;\ P(\omega_2|\mathbf{x}) = 0.736$
- $R(\alpha_1|\mathbf{x}) = 0 \cdot 0.264 + 0.736 = 0.736$,
  $R(\alpha_2|\mathbf{x}) = 0 \cdot 0.736 + 3 \cdot 0.264 = 0.792$.
- Don't delete the mail!

3. Bayes minimum-error-rate classifier

Minimum-Error-Rate Classification

Classification setting

- $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ (c possible states of nature)
- $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_c\}$ (α_i = decide ω_i, $1 \leq i \leq c$)

Zero-one (symmetrical) loss function

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

- Assign no loss (i.e. 0) to a correct decision
- Assign a unit loss (i.e. 1) to any incorrect decision (**equal cost**)

Minimum-Error-Rate Classification (Cont.)

$$\begin{aligned} R(\alpha_i \mid \mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) \cdot P(\omega_j \mid \mathbf{x}) \\ &= \sum_{j \neq i} \lambda(\alpha_i \mid \omega_j) \cdot P(\omega_j \mid \mathbf{x}) + \lambda(\alpha_i \mid \omega_i) \cdot P(\omega_i \mid \mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j \mid \mathbf{x}) \\ &= 1 - P(\omega_i \mid \mathbf{x}) \end{aligned}$$

error rate (误差率/错误率)
the probability that action α_i (decide ω_i) is wrong

Minimum error rate

Decide ω_i if $P(\omega_i \mid \mathbf{x}) > P(\omega_j \mid \mathbf{x})$ for all $j \neq i$