INSTITUT TEKNOLOGI BANDUNG

1920

Modul 3: Bayes Decision Theory

01 Bayes Decision Rule

Masayu Leylia Khodra
(masayu@informatika.org)

KK IF – Teknik Informatika – STEI ITB

Pengenalan Pola
(*Pattern Recognition*)

Bayes Decision Theory: What

Bayes decision theory is based on quantifying tradeoffs between various classification decision using probability and the cost that accompany such decision.

Assumption: decision problem posed in probabilistic terms and relevant probability values are known

Design classifiers to recommend decisions that minimize some total expected ”risk”.

Terminology

Labeled Dataset
$<\vec{x}, y>$

State of nature ω (random variable)
e.g., $\omega = y_1, \ldots, \omega = y_c$

Prior probability $P(\omega_i)$
e.g., prior knowledge of how likely is to get ω_i

Conditional probability density $P(x|\omega_i)$
e.g., likelihood pattern with feature value x given that the pattern belongs to class ω_i

Posterior cond. probability $P(\omega_i|x)$
e.g., probability that data belongs to class ω_i given feature value x

Bayes Decision Rule using Conditional Probability

One-dimensional feature x and binary class ω:

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

$$P(\omega_i|x) = \frac{P(x|\omega_i).P(\omega_i)}{P(x)} = \frac{likelihood * prior}{evidence}$$

Where $P(x) = \sum_{i=1}^{2} P(x|\omega_i).P(\omega_i)$

Decide ω_1 if $P(x|\omega_1).P(\omega_1) > P(x|\omega_2).P(\omega_2)$;
otherwise decide ω_2

https://www.cc.gatech.edu/~hic/CS7616/pdf/lecture2.pdf

Fish Classification

Image Dataset

Image	Label
	Seabass
	Seabass
	salmon

Random variable ω

ω_1 = sea bass (total=N_1)

ω_2 = salmon (total=N_2)

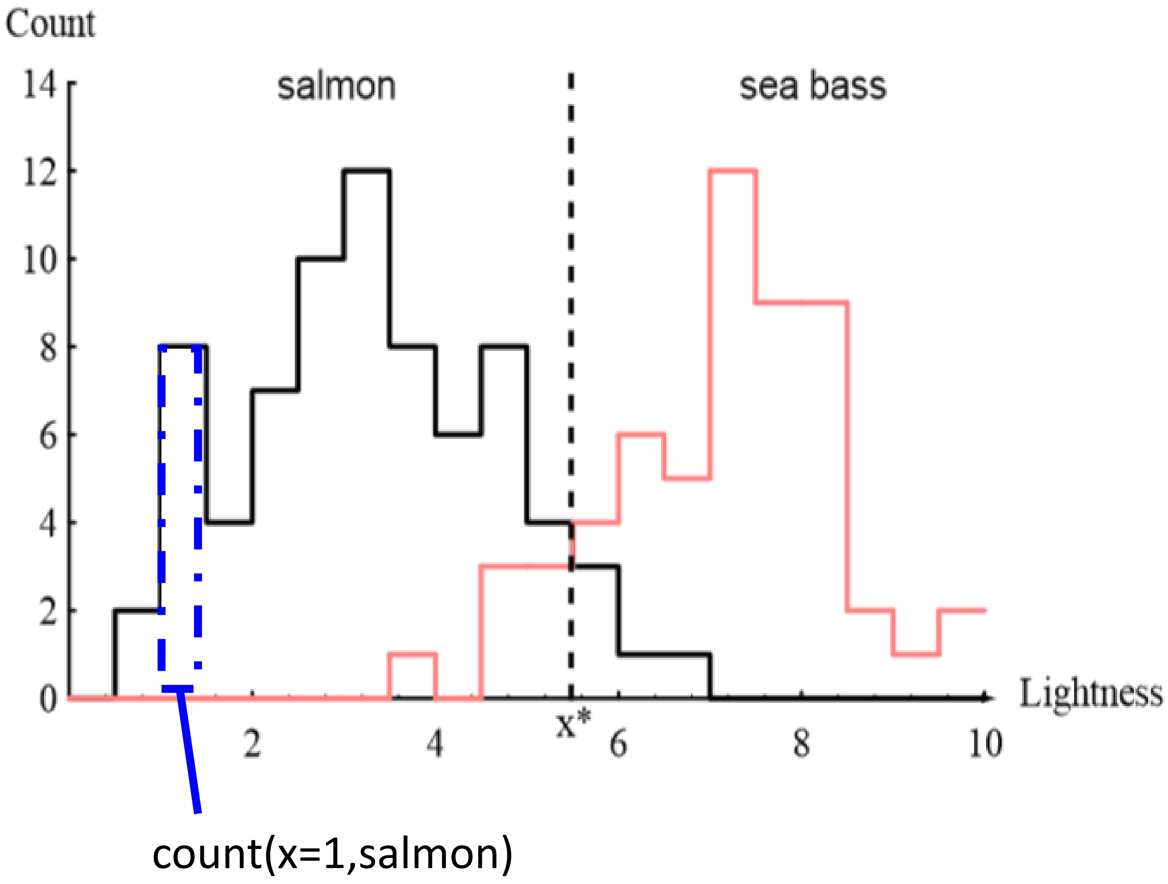
Prior probability: is the probability next fish observed is ω_i

$$P(\omega_1) = N_1 / (N_1 + N_2)$$

$$P(\omega_2) = N_2 / (N_1 + N_2)$$

https://www.thefishsociety.co.uk/fishopedia/sea-bass

Fish Classification: Feature Lightness



Conditional probability density

$P(x|\omega_i)=P(x,\omega_i)/P(\omega_i)$

$=count(x, \omega_i)/N_i$

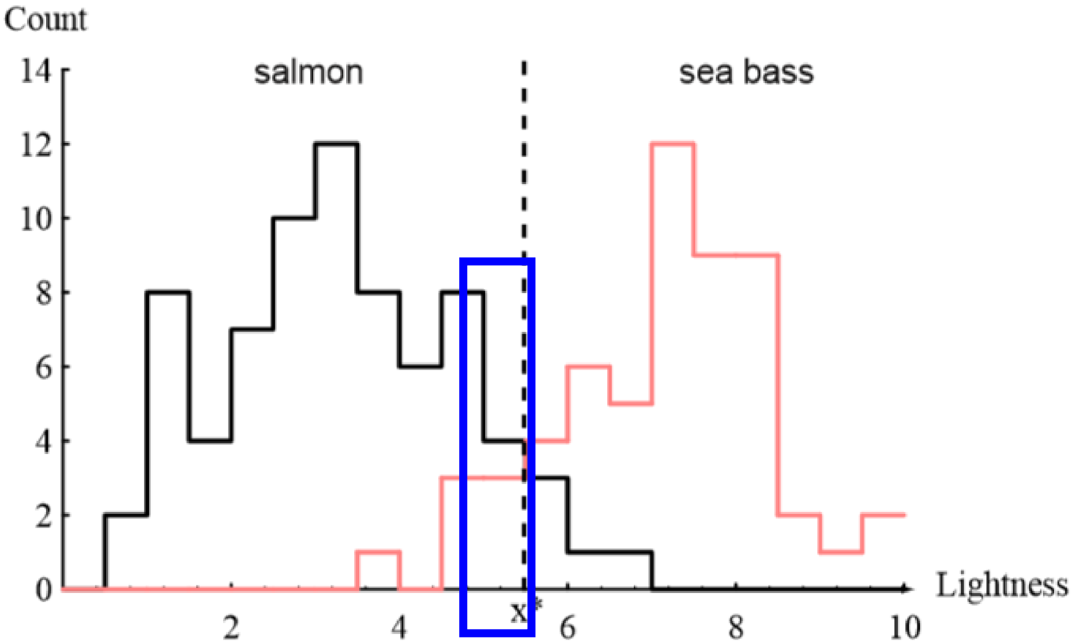
<u>New input x:</u>

Decide ω_1

if $P(x|\omega_1).P(\omega_1) > P(x|\omega_2).P(\omega_2)$;

otherwise decide ω_2

Fish Classification: Example



P(ω=salmon)=74/131=0.565
P(ω=sea bass)=57/131=0.435
P(x=5.0-5.5|ω=salmon)=4/74=0.054
P(x=5.0-5.5|ω=seabass)=3/57=0.053

- Input x = 5.0-5.5
- Posterior conditional probability: $P(\omega_i|x) \rightarrow P(x|\omega_i).P(\omega_i)$:

  P(ω=salmon|x = 5.0-5.5)
  =0.054*0.565=0.0305

  P(ω=sea bass|x = 5.0-5.5)
  =0.053*0.435=0.0230
- Decision: ω=salmon

Summary

Bayes Decision Theory

Bayes Decision Rule using Conditional Probability

Fish Classification

Minimum Risk Bayes Decision Rule

Modul 3: Bayes Decision Theory

02 Minimum Risk Bayes Decision Rule

Masayu Leylia Khodra
(masayu@informatika.org)

KK IF – Teknik Informatika – STEI ITB

Pengenalan Pola
(*Pattern Recognition*)

Action & Loss

- Allowing actions other than classification primarily allows the possibility of “rejection”.
- Example: a student wants to decide whether to take a course or not based on classification of the course.
  - ω={good, fair, bad}
  - Action={take, not_take}
- The loss function $\lambda(\alpha_i|\omega_j)$ specifies the cost of taking action α_i when correct classification category is ω_j.

https://adioshun.gitbooks.io/sensor_fusion/content/edx/edx-chapter2/2-4-bayesian-decision-theory.html

Conditional Risk (or Expected Loss)

- Example "student $\vec{\boldsymbol{x}}$ take the course ?":

	good	fair	bad
P(ω|$\vec{\boldsymbol{x}}$)	0.3	0.3	0.4

Loss	good	fair	bad
Take	0	5	30
Not_take	20	5	0

- Conditional risk or expected loss with taking action α_i is:

$$R(\alpha_i|\vec{\boldsymbol{x}}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\vec{\boldsymbol{x}})$$

- R(α=take|$\vec{\boldsymbol{x}}$)=0*0.3+5*0.3+30*0.4=13.5
- R(α=not_take|$\vec{\boldsymbol{x}}$)=20*0.3+5*0.3+0*0.4=7.5 → select not_take

https://adioshun.gitbooks.io/sensor_fusion/content/edx/edx-chapter2/2-4-bayesian-decision-theory.html

Minimum Risk Bayes Decision Rule

Allowing actions other than classification primarily allows the possibility of "rejection". General case with risks.

Computing expected loss $R(\alpha_i|\vec{x})$ for every action α_i

Choosing action α_i with minimum $R(\alpha_i|\vec{x})$

Two-category Classification

- $\omega=\{\omega_1, \omega_2\}$
- $\alpha=\{\text{decide } \omega_1, \text{decide } \omega_2\}$
- $\lambda_{ij}=\lambda(\alpha_i|\omega_j)$

$$R(\alpha_i|\vec{\boldsymbol{x}}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\vec{\boldsymbol{x}})$$

$$R(\alpha_1|\vec{\boldsymbol{x}}) = \sum_{j=1}^{2} \lambda(\alpha_1|\omega_j)P(\omega_j|\vec{\boldsymbol{x}})$$
$$= \lambda_{11}P(\omega_1|\vec{\boldsymbol{x}}) + \lambda_{12}P(\omega_2|\vec{\boldsymbol{x}})$$

$$R(\alpha_2|\vec{\boldsymbol{x}}) = \sum_{j=1}^{2} \lambda(\alpha_2|\omega_j)P(\omega_j|\vec{\boldsymbol{x}})$$
$$= \lambda_{21}P(\omega_1|\vec{\boldsymbol{x}}) + \lambda_{22}P(\omega_2|\vec{\boldsymbol{x}})$$

Bayes Decision Rule

n-dimensional feature $\vec{x}$ and binary class ω:

if $R(\alpha_1|\vec{x}) < R(\alpha_2|\vec{x})$: take action α_1 "decide ω_1";
otherwise take action α_2 "decide ω_2"

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

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

$$R(\alpha_1|\vec{x}) < R(\alpha_2|\vec{x})$$

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

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

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

if $\{(\lambda_{21}-\lambda_{11})P(\vec{x}|\omega_1)P(\omega_1)\} > \{(\lambda_{12}-\lambda_{22})P(\vec{x}|\omega_2)P(\omega_2)\}$:
take action α_1 "decide ω_1";
otherwise take action α_2 decide ω_2

Bayes Decision Rule (2)

n-dimensional feature $\vec{x}$ and binary class ω:

if $\{(\lambda_{21}-\lambda_{11})P(\vec{x}|\omega_1)P(\omega_1)\} > \{(\lambda_{12}-\lambda_{22})P(\vec{x}|\omega_2)P(\omega_2)\}$:
take action α_1 "decide ω_1";
otherwise take action α_2 decide ω_2

If $\underbrace{\frac{P(\vec{x}|\omega_1)}{P(\vec{x}|\omega_2)}}_{\text{Likelihood ratio}} > \underbrace{\frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)}}_{\text{threshold}}$: take action α_1 "decide ω_1";
otherwise decide ω_2

Bayes Decision Rule: Interpretation

If $\frac{P(\vec{\boldsymbol{x}}|\omega_1)}{P(\vec{\boldsymbol{x}}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)}$: take action α_1 "decide ω_1";

otherwise decide ω_2

If the likelihood ratio of class ω_1 and ω_2 exceeds a threshold value (independent of the input $\vec{\boldsymbol{x}}$), the optimal action is decide ω_1.

Example: Fish Classification

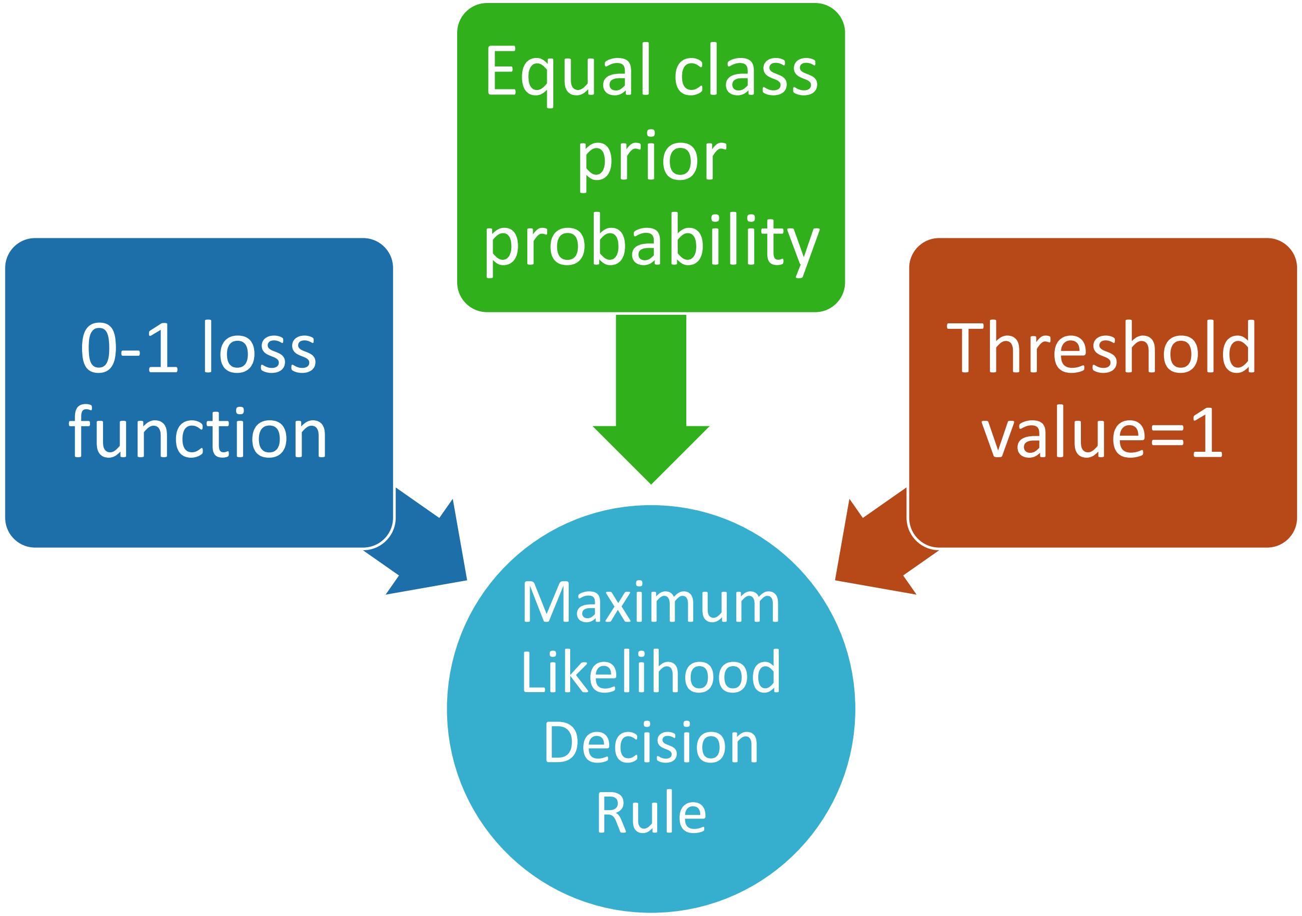
P(ω=salmon)=74/131=0.565
P(ω=sea bass)=57/131=0.435
P(x=5.0-5.5|ω=salmon)=4/74=0.054
P(x=5.0-5.5|ω=seabass)=3/57=0.053

Loss	Salmon	Sea bass
Salmon	λ_{11}=0	λ_{12}=1
Sea bass	λ_{21}=1	λ_{22}=0

- Likelihood ratio: LR=$\frac{P(\vec{x}|\omega_1)}{P(\vec{x}|\omega_2)} = \frac{0.054}{0.053} = 1.02$
- Threshold value: $\frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)} = \frac{(1-0)}{(1-0)}\frac{0.435}{0.565} = 0.77$
- LR > threshold → decide salmon

https://adioshun.gitbooks.io/sensor_fusion/content/edx/edx-chapter2/2-4-bayesian-decision-theory.html

Maximum Likelihood Decision Rule: Special Case



If $\frac{P(\vec{\boldsymbol{x}}|\omega_1)}{P(\vec{\boldsymbol{x}}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)}$:
take action α_1 "decide ω_1";
otherwise decide ω_2

Summary

Minimum risk bayes decision rule

Maximum likelihood decision rule

Classifier and Discriminant Function

Modul 3: Bayes Decision Theory

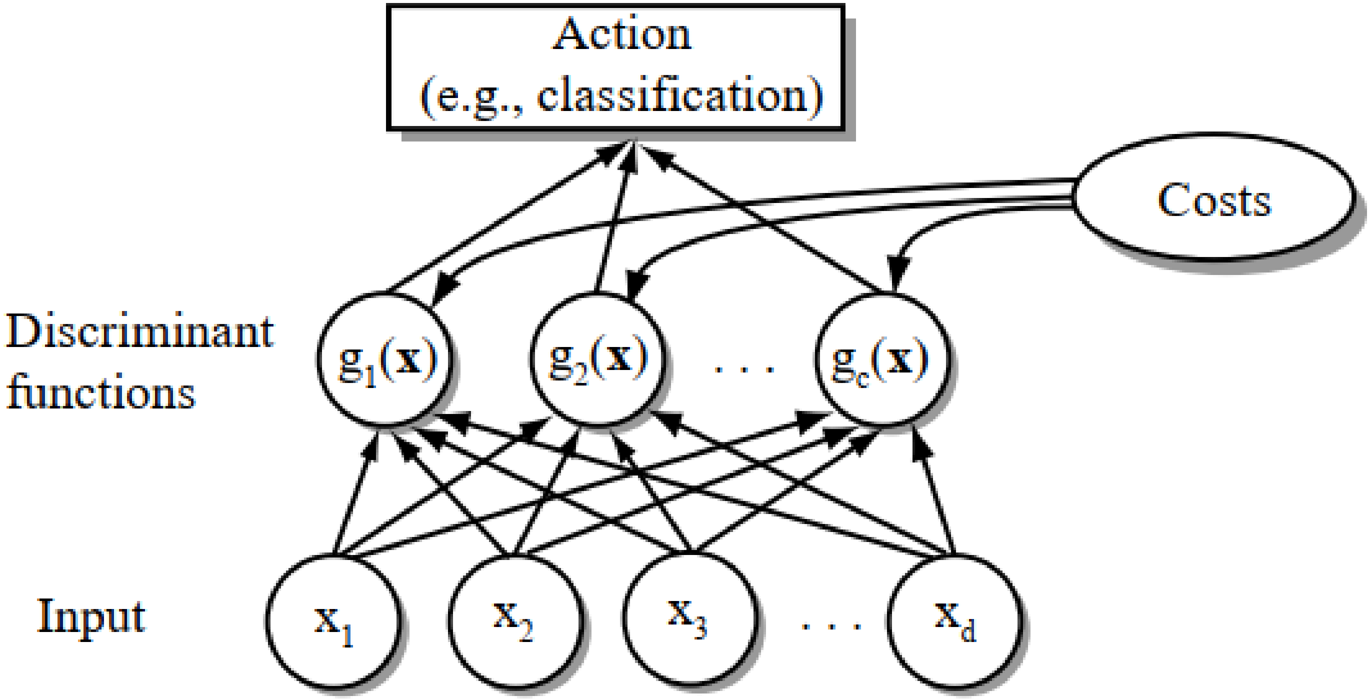
03 Classifier & Discriminant Function

Masayu Leylia Khodra
(masayu@informatika.org)

KK IF – Teknik Informatika – STEI ITB

Pengenalan Pola
(*Pattern Recognition*)

Classifier for Multicategory Case



Many different ways to represent classifiers or decision rules.
One of the most useful is in terms of “discriminant functions”

Classifier as Largest Discriminant

Discriminant functions $g_i(x)$, $i=1..c$

Dec. rule: Class ω_i if $g_i(x)>g_j(x)$ for all $j \neq i$

Maximum Discriminant Functions $g_i(x)$

$g_i(x) = -R(\alpha_i|\vec{x})$ for general case with risks

$g_i(x) = P(\omega_i|x)$: maximum posterior probability

$g_i(x) = \ln P(\vec{x}|\omega_i) + \ln P(\omega_i)$: simpler

Discriminant Functions for 2-Category

Assign ω_1 if $g(x) = g_1(x) - g_2(x) > 0$

$g(x) = P(\omega_1|x) - P(\omega_2|x)$

$g(x) = \ln\frac{P(\vec{x}|\omega_1)}{P(\vec{x}|\omega_2)} + \ln\frac{P(\omega_2)}{P(\omega_1)}$

EDUNEX ITB

Summary

Classifier: Maximum Discriminant Function

Discriminant Function for 2 Category

INSTITUT TEKNOLOGI BANDUNG

1920