Signal & Data Analysis in Neuroscience

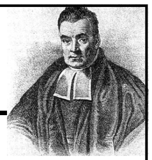
Bayesian Decisions

Izhar Bar-Gad

Room: 408 Phone: 7141 Email: izhar.bar-gad@biu.ac.il

1

Outline



- Bayesian decisions
- The Bayesian student
- The Bayesian doctor

Taken (almost) entirely from course:
Visual Recognition (236875) in the Technion

IBG

2

Decision theory

- Decision theory is an interdisciplinary area of study concerned with:

1. How decision-makers make decisions.
2. How optimal decisions can be reached.

- Decoding of neural information (and other types of encodings) relies heavily on decision theory.

IBG

3

Simple decision example

- Suppose that we know (via prior knowledge) that 25% of the newborns on April 1st are male and 75% are females.
- Our friend just had a newborn baby on that day but we forgot to ask about his/her gender. Should we buy the baby a pink or blue shirt?

(Yes, I know that colors don't matter but to this specific mother, they do)

- Thus, we need to guess the value of the variable *X* reflecting the state of nature using the a priori probabilities.

4

Decision error

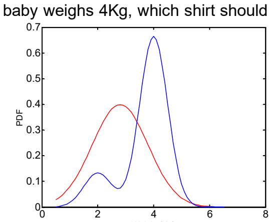
- Decision error → the probability of picking one possibility when the state of nature is different.
- The decision is done to minimize the error.
- In this example

$$P(error) = \begin{cases} \text{If we decide boy} \Rightarrow P(girl) \\ \text{If we decide girl} \Rightarrow P(boy) \end{cases}$$

5

Simple decision example – adding features

- Some features may give us information about the state of nature.
- Assuming that we know the weight distribution of boys (blue) and girls (red) and the happy mother told us that the baby weighs 4Kg, which shirt should we bring?



6

Simple decision example – conditional probability

- Assuming that the weight is represented by the random variable *Y*. The distribution of the weights assuming the gender is described by the class conditional probability p(y|x)
- So now the question becomes: what is that probability of a specific gender given the weight → p(x|y) ???

7

Conditional probability

- When 2 variables are statistically dependent, knowing the value of one of them lets us get a better estimate of the value of the other one.
- This is expressed by the conditional probability of *x* given *y*:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

- If x and y are statistically independent, then

$$P(x \mid y) = P(x)$$

8

Bayes' rule I

- The law of total probability: If event X can occur in m different ways $x_1, x_2, \dots, x_m$ and if they are mutually exclusive → the probability of X is the sum of the probabilities $x_1, x_2, \dots, x_m$.

$$P(y) = \sum_x P(x, y).$$

- From the definition of conditional probability

$$P(x, y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

9

Bayes' rule II

$$P(x\mid y)=\frac{P(y\mid x)P(x)}{P(y)}$$

$$P(x\mid y)=\frac{P(y\mid x)P(x)}{\sum_x P(x,y)}=\frac{P(y\mid x)P(x)}{\sum_x P(y\mid x)P(x)}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

IBG

10

Bayes' rule – continuous case

- For continuous random variable we refer to densities rather than probabilities; in particular,

$$p(x\mid y)=\frac{p(x,y)}{p(y)}$$

- The Bayes' rule for densities becomes:

$$p(x\mid y)=\frac{p(y\mid x)p(x)}{\int_{-\infty}^{\infty} p(y\mid x)p(x)dx}$$

IBG

11

Bayes' rule - importance

- *x* is termed the cause & y is termed the effect. Assuming *x* is present, we know the likelihood of *y* to be observed
- Bayes' rule allows to determine the likelihood of a cause *x* given an observation *y*. Note: there may be many causes producing *y*.
- Bayes' rule shows how probability for *x* changes from **prior** p(*x*) before we observe anything, to **posterior** p(x| y) once we have observed *y*.

IBG

12

Bayes' decision rule

- **Decision:**

  boy : if P(boy|weight) > P(girl|weight)
  girl : otherwise

or

  boy : if P (weight |boy)P(boy) > P(weight|girl)P(girl)
  girl : otherwise

- **Error:**

$$P(error \mid weight) = \begin{cases} \text{If we decide boy} \Rightarrow P(girl \mid weight) \\ \text{If we decide girl} \Rightarrow P(boy \mid weight) \end{cases}$$

P(error|weight) = min [P(boy|weight) , P(girl|weight)]

IBG

13

Loss function

- The problem arises when different decisions have different consequences (for example: pink shirt for a boy is less acceptable in many cultures than a blue one for a girl).
- Loss (or cost) function states exactly how costly each action is, and is used to convert a probability determination into a decision. Loss functions let us treat situations in which some kinds of classification mistakes are more costly than others.

IBG

14

Expected loss

- Suppose that we observe a particular ***y*** and that we contemplate taking action α_i .
- If the true state of nature is x_j the loss is $\lambda(\alpha_i \mid x_j)$
- Before we have done an observation the expected loss is $R(\alpha_i) = \sum_{j=1}^{C} \lambda(\alpha_i \mid x_j) P(x_j)$
- After the observation the expected risk which is called now the conditional risk is given by

$$R(\alpha_i \mid y) = \sum_{j=1}^{C} \lambda(\alpha_i \mid x_j) P(x_j \mid y)$$

IBG

15

Bayes' decision rule

- Compute the conditional risk for each action

$$R(\alpha_i \mid y) = \sum_{j=1}^{C} \lambda(\alpha_i \mid x_j) P(x_j \mid y)$$

- Select the action α_i for which $R(\alpha_i \mid y)$ is minimal.
- The resulting minimum risk is called the Bayes Risk, denoted R*, and is the best performance that can be achieved.

IBG

16

Optimal Bayes Decision Strategies

- A ***strategy*** or ***decision function*** $\alpha(y)$ is a mapping from observations to actions.
- The ***total risk*** of a decision function is given by

$$E_{p(y)}[R(\alpha(y) \mid y)] = \sum_{y} p(y) \cdot R(\alpha(y) \mid y)$$

- A decision function is *optimal* if it minimizes the total risk. This optimal total risk is called *Bayes risk*.

IBG

17

Outline

- Bayesian decisions
- The Bayesian student
- The Bayesian doctor

IBG

18

The student dilemma

- A student needs to achieve a decision on which courses to take, based only on his first lecture.
- From his previous experience, he knows the **prior probabilities** :

Quality of the course	good	fair	bad
$P(x_j)$ → prior	0.2	0.4	0.4

IBG

19

The student dilemma

- The student also knows the **class-conditionals**:

$P(y\|x_j)$	good	fair	bad
Interesting lecture	0.8	0.5	0.1
Boring lecture	0.2	0.5	0.9

- The **loss function** is given by the matrix

$\lambda(a_i\|x_j)$	good course	fair course	bad course
Taking the course	0	5	10
Not taking the course	20	5	0

IBG

20

The student dilemma

- The student wants to make an optimal decision of taking the course based on the first lecture.
- The probability of hearing an interesting lecture:

P(interesting)= P(interesting|good)* P(good)
+ P(interesting|fair)* P(fair)
+ P(interesting|bad)* P(bad)
= 0.8*0.2+0.5*0.4+0.1*0.4 = 0.4

P(boring)= 1-P(interesting) = 1-0.4 =0.6

- Assuming that the lecture was interesting, what are the **posterior** probabilities of each of the 3 possible "states of nature"?

IBG

21

The student dilemma

$$P(\text{good course}|\text{interesting lecture}) = \frac{P(\text{interesting}|\text{good})\Pr(\text{good})}{P(\text{interesting})} = \frac{0.8*0.2}{0.4} = 0.4$$

$$P(\text{fair}|\text{interesting}) = \frac{P(\text{interesting}|\text{fair})P(\text{fair})}{P(\text{interesting})} = \frac{0.5*0.4}{0.4} = 0.5$$

- We can get P(bad|interesting)=0.1 either by the same method, or by noting that it complements to 1 the above two.

22

The student dilemma

- The student needs to minimize the conditional risk.

$$R(\alpha_i \mid y) = \sum_{j=1}^{c} \lambda(\alpha_i \mid x_j) P(x_j \mid y)$$

- In this case there are only two possible actions: taking or not taking the course.

R(taking|interesting)= P(good|interesting)λ(taking course|good)
+P(fair|interesting)λ(taking course|fair)
+P(bad|interesting)λ(taking course|bad)
=0.4*0+0.5*5+0.1*10=3.5

R(not taking|interesting)=P(good|interesting)λ(not taking course|good)
+P(fair|interesting)λ(not taking course|fair)
+P(bad|interesting)λ(not taking course|bad)
=0.4*20+0.5*5+0.1*0=10.5

23

The student dilemma

- So, if the first lecture was interesting, the student will minimize the conditional risk <u>by taking the course</u>.
- In order to construct the full decision function, we need to define the risk minimization action for the case of boring lecture, as well.

24

Outline

- Bayesian decisions
- The Bayesian student
- The Bayesian doctor

IBG

25

The Bayesian Doctor Example

A person doesn't feel well and goes to the doctor.
Assume two states of nature:
x_1 : The person has a common flu.
x_2 : The person has a vicious bacterial infection.

The doctors *prior* is: $p(x_1) = 0.9 \quad p(x_2) = 0.1$

This doctor has two possible actions:
a_1 = Prescribe hot tea.
a_2 = Prescribe antibiotics.

The doctor can use prior and predict optimally: always flu.

Therefore doctor will always prescribe hot tea.

IBG

26

The Bayesian Doctor Example

- But there is very high risk: Although this doctor can diagnose with very high rate of success using the prior, (s)he can lose a patient once in a while.
- Denote the two possible actions:
  a_1 = prescribe hot tea
  a_2 = prescribe antibiotics
- Now assume the following cost (loss) matrix:

$\lambda_{i,j} =$

	x_1	x_2
a_1	0	10
a_2	1	0

IBG

27

The Bayesian Doctor Example

- Choosing a_1 results in ***expected risk*** of

$$R(a_1) = p(x_1)\cdot\lambda_{1,1} + p(x_2)\cdot\lambda_{1,2}$$
$$= 0 + 0.1\cdot 10 = 1$$

- Choosing a_2 results in expected risk of

$$R(a_2) = p(x_1)\cdot\lambda_{2,1} + p(x_2)\cdot\lambda_{2,2}$$
$$= 0.9\cdot 1 + 0 = 0.9$$

- So, considering the costs it's much better (and optimal!) to always give antibiotics.

28

The Bayesian Doctor Example

- However, doctors can also produce some *observations such* as performing a blood test.
- The possible results of the blood test are:
  y_1 = negative (no bacterial infection)
  y_2 = positive (infection)
- Blood tests are never conclusive leading to the ***class conditional*** probabilities.

$$p(y_1 \mid x_1) = 0.8 \qquad p(y_2 \mid x_1) = 0.2$$
$$p(y_1 \mid x_2) = 0.3 \qquad p(y_2 \mid x_2) = 0.7$$

29

The Bayesian Doctor Example

- Define the conditional risk given the observation $R(a_i \mid y) = \sum_{\omega_j} p(x_j \mid y)\cdot\lambda_{i,j}$
- We would like to compute the conditional risk for each action and observation so that the doctor can choose an optimal action that minimizes risk.
- How can we compute $p(x_j \mid y)$?
- We use the class conditional probabilities and Bayes inversion rule.

30

The Bayesian Doctor Example

- The results of the blood test follow the probabilities:

$$p(y_1) = p(y_1 \mid x_1)\cdot p(x_1) + p(y_1 \mid x_2)\cdot p(x_2)$$
$$= 0.8\cdot 0.9 + 0.3\cdot 0.1$$
$$= 0.75$$
$$p(y_2) = 1 - p(y_1) = 0.25$$

31

The Bayesian Doctor Example

$$\mathrm{R}(a_1 \mid y_1) = p(x_1 \mid y_1)\cdot \lambda_{1,1} + p(x_2 \mid y_1)\cdot \lambda_{1,2}$$
$$= 0 + p(x_2 \mid y_1)\cdot 10$$
$$= 10\cdot \frac{p(y_1 \mid x_2)\cdot p(x_2)}{p(y_1)}$$
$$= 10\cdot \frac{0.3\cdot 0.1}{0.75} = 0.4$$

$$R(a_2 \mid y_1) = p(x_1 \mid y_1)\cdot \lambda_{2,1} + p(x_2 \mid y_1)\cdot \lambda_{2,2}$$
$$= p(x_1 \mid y_1)\cdot 1 + p(x_2 \mid y_1)\cdot 0$$
$$= \frac{p(y_1 \mid x_1)\cdot p(x_1)}{p(y_1)}$$
$$= \frac{0.8\cdot 0.9}{0.75} = 0.96$$

32

The Bayesian Doctor Example

$$R(a_1 \mid y_2) = p(x_1 \mid y_2)\cdot \lambda_{1,1} + p(x_2 \mid y_2)\cdot \lambda_{1,2}$$
$$= 0 + p(x_2 \mid y_2)\cdot 10$$
$$= 10\cdot \frac{p(y_2 \mid x_2)\cdot p(x_2)}{p(y_2)}$$
$$= 10\cdot \frac{0.7\cdot 0.1}{0.25} = 2.8$$

$$R(a_2 \mid y_2) = p(x_1 \mid y_2)\cdot \lambda_{2,1} + p(x_2 \mid y_2)\cdot \lambda_{2,2}$$
$$= p(x_1 \mid y_2)\cdot 1 + p(x_2 \mid y_2)\cdot 0$$
$$= \frac{p(y_2 \mid x_1)\cdot p(x_1)}{p(y_2)}$$
$$= \frac{0.2\cdot 0.9}{0.25} = 0.72$$

33

The Bayesian Doctor Example

- To summarize:
$$R(a_1 \mid y_1) = 0.4$$
$$R(a_2 \mid y_1) = 0.96$$
$$R(a_1 \mid y_2) = 2.8$$
$$R(a_2 \mid y_2) = 0.72$$
- Given an observation *y*, we can minimize the expected loss by minimizing the conditional risk.
- The doctor chooses:
  - Hot tea if blood test is negative
  - Antibiotics otherwise.

IBG

34

Optimal Bayes Decision Strategies

- The ***total risk*** of a decision function is given by
$$E_{p(y)}[R(\alpha(y) \mid y)] = \sum_y p(y) \cdot R(\alpha(y) \mid y)$$
- A decision function is *optimal* if it minimizes the total risk. This optimal total risk is called *Bayes risk*.
- In the Bayesian doctor example:
  - The prior risk (the doctor always gives antibiotics): 0.9
  - The Bayes risk: 0.75*0.4+0.25*0.72=0.48

IBG

35