# Baye’s Theorem

# Decision theory

- Decision theory is the study of making decisions that have a significant impact

- Decision-making is distinguished into:
  - Decision-making under certainty
  - Decision-making under non-certainty
    - Decision-making under risk
    - Decision-making under uncertainty

# Probability theory

- Most decisions have to be taken in the presence of uncertainty
- Basic elements of probability theory:
  - Random variables describe aspects of the world whose state is initially unknown
  - Each random variable has a domain of values that it can take on (discrete, continuous)
  - An atomic event is a complete specification of the state of the world

# Probability Theory

- All probabilities are between 0 and 1
- The sum of probabilities for the atomic events of a probability space must sum up to 1

# Prior

- **Priori Probabilities** or Prior reflects our prior knowledge of how likely an event occurs.

# **Class Conditional probability (**posterior)

- When we have information concerning previously unknown random variables then we use **posterior** or conditional probabilities: P(a/b) the **probability of a given event a that we know b**

$$P(a/b) = \frac{P(a \wedge b)}{P(b)}$$

- Alternatively this can be written (the product rule):

  $P(a \wedge b) = P(a/b)P(b)$

# Baye's rule

- The product rule can be written as:
- $P(a \wedge b) = P(a|b)P(b)$
- $P(a \wedge b) = P(b|a)P(a)$

- By equating the right-hand sides:

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}$$

- This is known as Baye's rule

# Posterior Probabilities

- Define $p(c_j/x)$ as the posteriori probability
- We use Baye's formula to convert the prior to posterior probability

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

# Bayes Classifiers

- Bayesian classifiers use **Bayes theorem**, which says

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

- $\boldsymbol{p(c_j \mid x)}$ = probability of instance x being in class $c_j$,
  This is what we are trying to compute
- $\boldsymbol{p(x \mid c_j)}$ = probability of generating instance *x* given class $c_j$,
  We can imagine that being in class $c_j$ causes you to have feature x with some probability
- $\boldsymbol{p(c_j)}$ = probability of occurrence of class $c_j$,
  This is just how frequent the class $c_j$ is in our database
- $\boldsymbol{p(x)}$ = probability of instance *x* occurring

  This can actually be ignored since it is the same for all classes

# Bayes Formula

- Suppose the priors $P(c_j)$ and conditional densities $p(x|c_j)$ are known,

*prior* *likelihood*

*posterior*

$$\underbrace{P(C \mid \mathbf{x})}_{\textit{posterior}} = \frac{\overbrace{P(C)}^{\textit{prior}}\,\overbrace{p(\mathbf{x} \mid C)}^{\textit{likelihood}}}{\underbrace{p(\mathbf{x})}_{\textit{evidence}}}$$

*evidence*

# Bayesian Decision Theory

Tradeoffs between various decisions using probabilities and costs that accompany such decisions.

Example: **Patient has trouble breathing**

– **Decision**: Asthma versus Lung cancer

1. Decide lung cancer when person has asthma
   - Cost: moderately high (e.g., order unnecessary tests, scare patient)
2. Decide asthma when person has lung cancer
   - **Cost: very high (e.g., lose opportunity to treat cancer at early stage, death)**

# Decision Rules

Progression of decision rules:

1. Decide based on prior probabilities
2. Decide based on posterior probabilities
3. Decide based on risk

# Fish Sorting Example

- C ➔ class

  C=c1 (sea bass)

  C=c2 (salmon)
- P(c1) is the prior probability that the next fish is a sea bass
- P(c2) is the prior probability that the next fish is a salmon

# Decision based on prior probabilities

- Assume $P(c1) + P(c2) = 1$
- Decision ??
- Decide ➔
  - C1 if $P(c1) > P(c2)$
  - C2 otherwise
- Error probability
  $p(error) = \min(P(c1), P(c2))$

# Decision based on class conditional probabilities

- Let x be a continuous random variable
- Define $p(x/cj)$ as the conditional probability density ($j=1,2$)
- $P(x/c1)$ and $P(x/c2)$ describe the difference in measurement between populations of sea bass and Solomon

# Making a Decision

- Decision ??? (After observing x value)
- Decide :
  - C1 if $P(c1/x) > P(c2/x)$
  - C2 otherwise
- $P(c1/x) + P(c2/x)=1$

# Probability of Error

- $P(error/x)$ :
  - $P(c1/x)$ if we decide c2
  - $P(c2/x)$ if we decide c1
- $P(error/x) = \min \{P(c1/x), P(c2/x)\}$

Assume that we have two classes

$c_1$ = male, and $c_2$ = female.

We have a person whose sex we do not know, say "**gagan**" or g.

Classifying **gagan** as male or female is equivalent to asking is it more probable that **gagan** is male or female, I.e which is greater $p$(male | gagan) or $p$(female | gagan)

What is the probability of being called "gagan" given that you are a male?

What is the probability of being a male?

$$p(\text{male} \mid \text{gagan}) = \frac{p(\text{gagan} \mid \text{male})\, p(\text{male})}{p(\text{gagan})}$$

What is the probability of being named "gagan"?

# Dataset

| **Name** | **Sex** |
|---|---|
| Gagan | Male |
| Namita | Female |
| Gagan | Female |
| Gagan | Female |
| Ram | Male |
| Sunita | Female |
| Jamuna | Female |
| Ram | Male |

$$p(c_j \mid g) = p(\mathrm{g} \mid \mathrm{c}_j)\, p(c_j) \;/\; p(g)$$

| Name | Sex |
|---|---|
| Gagan | Male |
| Namita | Female |
| Gagan | Female |
| Gagan | Female |
| Ram | Male |
| Sunita | Female |
| Jamuna | Female |
| Ram | Male |

$$p(c_j \mid g) = \frac{p(g \mid c_j)\, p(c_j)}{p(g)}$$

**Gagan**

$$p(\text{male} \mid gagan) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8}$$

$$p(\text{female} \mid gagan) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8}$$

Gagan is more likely to be a Female.

# Gagan is a female

# Advantages/Disadvantages of Naïve Bayes

- Advantages
  - Fast to train (single scan). Fast to classify
  - Handles real and discrete data
  - Handles streaming data well
- Disadvantages
  - Assumes independence of features