Algoritmos y Programación III

Andrés A. Aristizábal P. aaaristizabal@icesi.edu.co
Angela Villota

Departamento de Computación y Sistemas Inteligentes

UNIVERSIDAD ICESI

2024-2

Agenda

1. **Reasoning with uncertainty**
   - Bayes Network Example: The Monty Hall Problem
   - Maximum Likelihood Estimation
   - Exercises

Bayes Network Example: The Monty Hall Problem

The Monty What?

The Monty What?

- Context
- Development

Agenda

1. **Reasoning with uncertainty**
   - Bayes Network Example: The Monty Hall Problem
   - Maximum Likelihood Estimation
   - Exercises

Maximum Likelihood Estimation

What is MLE?

Maximum Likelihood Estimation

What is MLE?

- Maximum Likelihood Estimation (MLE) is a method used in statistics to estimate the parameters of a statistical model.
- The basic idea behind MLE is to find the parameter values that maximize the likelihood function, which measures how likely the observed data are under the given statistical model.

How is MLE used with respect to Bayesian networks?

Maximum Likelihood Estimation

What is MLE?

- Maximum Likelihood Estimation (MLE) is a method used in statistics to estimate the parameters of a statistical model.
- The basic idea behind MLE is to find the parameter values that maximize the likelihood function, which measures how likely the observed data are under the given statistical model.

How is MLE used with respect to Bayesian networks?

- MLE is used to estimate the conditional probability tables (CPTs) associated with each node in the network.

Maximum Likelihood Estimation

In the context of Bayesian networks, MLE is used to estimate the conditional probabilities associated with each node in the network, given data.

Steps:

1. **Define the Network Structure:** The Bayesian network?s structure is typically known, consisting of nodes (variables) and directed edges (dependencies).
2. **Collect Data:** Gather a dataset where each instance consists of values for all the variables in the network.
3. **Estimate Parameters (Conditional Probabilities):** For each node X with parent nodes $\text{Parents}(X)$, the goal is to estimate the conditional probability $P(X \mid \text{Parents}(X))$. MLE helps find the conditional probability distribution that maximizes the likelihood of the observed data.

Example

Imagine a simple Bayesian network with nodes A (disease) and B (symptom), where $A \rightarrow B$ means the disease causes the symptom.
Given a dataset containing the information (whether patients have the disease and show the symptom), MLE can be used to estimate the probabilities:

- $P(A)$ (the probability of having the disease).
- $P(B \mid A)$ (the probability of showing symptoms, given the disease).

If we observe a dataset of patients, MLE will determine the values of $P(A)$ and $P(B \mid A)$ that best explain the data.

Maximum Likelihood Estimation

What is the fundamental concept behind MLE?

Maximum Likelihood Estimation

What is the fundamental concept behind MLE?

- To select the parameters that make the observed data most likely.
- MLE is used for the data having n independent and identically distributed samples: $X_1, X_2, X_3, ..., X_n$.

Maximum Likelihood Estimation

First, what is Likelihood, and how it differs from Probability?

Maximum Likelihood Estimation

First, what is Likelihood, and how it differs from Probability?

- **Probability** refers to the chance of an event occurring, given some model or known parameters.
- **Likelihood** works in the reverse direction: it tells us how likely a set of parameters (or a model) is, given observed data.
- Likelihood is about inference: what parameter value best explains the observed outcome?
- When calculating the probability of a given outcome, you assume the model's parameters are reliable.
- However, when you calculate the likelihood, you're attempting to determine whether the parameters in a model can be trusted based on the sample data you have observed.

Maximum Likelihood Estimation

Let us look at an example

Maximum Likelihood Estimation

Let us look at an example

Probability

Suppose we have a fair coin, meaning the probability of heads $P(H)$ is 0.5, and the probability of tails $P(T)$ is 0.5. The probability of getting heads in a single toss is:

$$P(H) = 0{,}5$$

In this case, the parameters (the fairness of the coin) are known, and we calculate the probability of possible outcomes (heads or tails).

Maximum Likelihood Estimation

Let us look at an example

Probability

Suppose we have a fair coin, meaning the probability of heads $P(H)$ is 0.5, and the probability of tails $P(T)$ is 0.5. The probability of getting heads in a single toss is:

$$P(H) = 0{,}5$$

In this case, the parameters (the fairness of the coin) are known, and we calculate the probability of possible outcomes (heads or tails).

Likelihood

We don't know if the coin is fair. Then, we toss the coin 10 times and observe 7 heads and 3 tails.

- Now, suppose we hypothesize that $p = 0{,}7$ (a biased coin). The **likelihood** of observing 7 heads and 3 tails is higher for this value than for $p = 0{,}5$, since it aligns better with the observed data.

Thus, **likelihood** allows us to estimate which value of p best explains the observed data (7 heads, 3 tails).

Maximum Likelihood Estimation

Which are the types of likelihood functions?

Maximum Likelihood Estimation

Which are the types of likelihood functions?

- Discrete probability distribution:
  - $\mathcal{L}(\theta \mid x) = p_\theta(x) = P_\theta(X = x) = P(X = x \mid \theta)$
  - Let X be a discrete random variable with probability mass function p depending on a parameter θ
  - The likelihood is the probability that a particular outcome x is observed when the true value of the parameter is θ, equivalent to the probability mass on x
- Continuous probability distribution:
  - $\mathcal{L}(\theta \mid x) = f_\theta(x)$
  - Let X be a random variable following an absolutely continuous probability distribution with density function f (a function of x) which depends on a parameter θ

Maximum Likelihood Estimation

What is MLE?

Maximum Likelihood Estimation

What is MLE?

- Maximum Likelihood Estimation (MLE) is a method used in statistics to estimate the parameters of a statistical model.
- The basic idea behind MLE is to find the parameter values that maximize the likelihood function, which measures how likely the observed data are under the given statistical model.

Maximum Likelihood Estimation

How does MLE work?

Maximum Likelihood Estimation

How does MLE work?

1. **Define the Statistical Model:** First, you need to specify a statistical model that describes the distribution of your data. This model typically depends on one or more parameters that you want to estimate.
2. **Define the Likelihood Function:** The likelihood function measures how likely the observed data are under the given statistical model. It is a function of the parameters of the model.
3. **Maximize the Likelihood Function:** The goal of MLE is to find the values of the parameters that maximize the likelihood function. This can often be done analytically by taking derivatives of the likelihood function with respect to the parameters and setting them equal to zero. In some cases, numerical optimization techniques may be necessary to find the maximum.
4. **Interpret the Results:** Once you have found the maximum likelihood estimates of the parameters, you can use them to make inferences about the population or to make predictions about future observations.

Let us look at an example

Let us look at an example

Example

Coin flipping

- *Imagine you're trying to estimate the probability of getting heads (success) when flipping a biased coin.*
- *You're given a coin and asked to flip it 10 times, resulting in 7 heads and 3 tails.*
- *You want to use MLE to estimate the probability of getting heads.*

Solution

Define the Statistical Model:

Solution

Define the Statistical Model:

- *We'll use the binomial distribution to model the number of heads obtained in 10 coin flips.*
- *The binomial distribution has two parameters: n (the number of trials) and p (the probability of success on each trial).*

Solution

Define the Likelihood function:

Solution

Define the Likelihood function:

- *The likelihood function measures how likely it is to observe the given data (7 heads and 3 tails) for different values of the parameter p in our case.*
- *Having* $\mathcal{L}(p \mid n, k) = (\binom{n}{k} p^k (1 - p^{n-k}))$
- *The likelihood function for the binomial distribution is given by:*

$$\mathcal{L}(p \mid 10, 7) = (\binom{10}{7} p^7 (1 - p^3))$$

Solution

Maximize the Likelihood Function:

Maximum Likelihood Estimation

Solution

Maximize the Likelihood Function:

- *To find the value of p that maximizes the likelihood function, we can take the derivative of the likelihood function with respect to p, set it equal to zero, and solve for p.*

$$\frac{d}{dp}\mathcal{L}(p \mid 10, 7) = 0)$$

- *Solving this equation will give us the maximum likelihood estimate of p.*

Solution

Interpret the results:

Maximum Likelihood Estimation

Solution

Interpret the results:

- *Once we find the maximum likelihood estimate of p, we can interpret it as the best estimate of the probability of getting heads when flipping this coin.*

Maximum Likelihood Estimation

Solution

Finally:

Solution

Finally:

- *In this example, given that you observed 7 heads out of 10 flips, the maximum likelihood estimate of the probability of getting heads (p) would be the value of p that maximizes the likelihood function* $\mathcal{L}(p \mid 10, 7)$
- *After solving the equation and finding the value of p, let's say we find that p=0.7.*
- *This means that, based on the observed data, the best estimate for the probability of getting heads when flipping this coin is 0.7, or 70 %.*

Maximum Likelihood Estimation

How is MLE used with respect to Bayesian networks?

Maximum Likelihood Estimation

How is MLE used with respect to Bayesian networks?

- MLE is used to estimate the conditional probability tables (CPTs) associated with each node in the network.
- For this we need to obtain a dataset where the variables represented in the Bayesian network are observed.
- This dataset will be used to estimate the parameters of the CPTs.
- For discrete variables, this involves counting the frequencies of different outcomes in the dataset and normalizing to obtain probabilities.

Maximum Likelihood Estimation

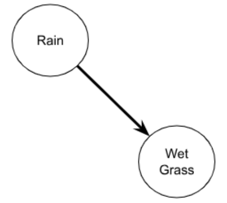
Let us look at an example

Maximum Likelihood Estimation

Let us look at an example

Example

Consider the following Bayesian network



- *Rain is a binary variable indicating whether it is raining or not.*
- *Wet Grass is a binary variable indicating whether the grass is wet or not.*
- *Suppose we have collected data on whether it was raining and whether the grass was wet for several days.*
- *We want to estimate the conditional probability table (CPT) for the Wet Grass node given the Rain node.*

Let's say our dataset looks like this:

Example

Day	*Rain*	*Wet Grass*
1	*0*	*0*
2	*1*	*1*
3	*1*	*1*
4	*0*	*1*
5	*1*	*1*

Maximum Likelihood Estimation

Solution

To estimate the parameters of the CPT for Wet Grass given Rain using MLE, we count the frequencies of each combination of values:

- *P(WetGrass=1 | Rain = 0) = 1/2*
- *P(WetGrass=1 | Rain = 1) = 3/3 = 1*
- *P(WetGrass=0 | Rain = 0) = 1/2*
- *P(WetGrass=0 | Rain = 1) = 0/3 = 0*

These probabilities represent our estimates of the conditional probabilities based on the observed data and then we can use them for inference tasks within the Bayesian network.

Maximum Likelihood Estimation

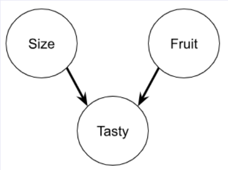
Let us look at another example

Maximum Likelihood Estimation

Let us look at another example

Example

Consider the following Bayesian network



- *Fruit is a binary variable indicating whether it is an apple or a banana.*
- *Size is another binary variable which shows if the fruit is small or large.*
- *Finally, Tasty is also a binary variable that indicates if a fruit is tasty or not.*
- *Let's fin the probability of a fruit being tasty.*

Agenda

1. **Reasoning with uncertainty**
   - Bayes Network Example: The Monty Hall Problem
   - Maximum Likelihood Estimation
   - Exercises

Exercises[1]

Exercise

Implement the Bayesian Network for the Alarm problem.

Exercises[2]

Exercise

Given this data:

	outlook	temperature	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Use bayesian networks to infer the probabilities:

1. *Of a tennis game being played if the temperature is hot, humidity high and outlook sunny*
2. *Of a tennis game being played if the temperature is mild, humidity normal, it is not windy and the outlook is sunny*
3. *Of a tennis game being played if the temperature is cool and the outlook is rainy*