Overview Probability theory Conditional Probability

Probabilistic Graphical Models in Computer Vision (IN2329)

Csaba Domokos
Summer Semester 2017

1. Introduction

Agenda for today’s lecture

IN2329 - Probabilistic Graphical Models in Computer Vision 1. Introduction – 8 / 39

1. Administrative
2. Overview of the course
3. Introduction to Probability Theory
   - Basic definitions
   - Conditional probability, Bayes’ rule
   - Independence, conditional independence

The course: IN2329

The course Probabilistic Graphical Models in Computer Vision will be organized as follows:
- Lectures will take place between 10:00–12:00 on Mondays in Room 02.09.023.
- Tutorials will take place between 14:00–16:00 on Mondays in Room 02.05.014.

Contact details of the lecturer:
Name: Dr. Csaba Domokos
Email: csaba.domokos@in.tum.de
Office: 02.09.060
Feel free to contact me!

Tutorials

Participation at the tutorial is not mandatory, but highly recommended.
The tutorials combine theoretical and programming assignments:
- Theoretical assignments will help to understand the topics of the lecture.
- Programming assignments will help to apply the theory to practical problems.
- Assignment distribution:
  between 11.00–11.15 on Mondays in Room 02.09.023.
- Theoretical assignment due:
  between 11.00–11.15 on Mondays in Room 02.09.023.
- Assignment presentation:
  between 14.00–16.00 on Mondays in Room 02.05.014.

To promote team work, please form groups of two or three students in order to
solve and submit the assignments.
Active students who solve 60% of the assignments earn a bonus.

Exam

- The exam will be oral (max. 30 mins.).
- The exam will take place 7th–9th of August.
- Students need to be registered prior to the exam:
  May, 29th–June, 30th via TUM online.

To achieve the bonus, the following requirements have to be fulfilled:
- Theory:
  - 60% of all theoretical assignments have to be solved.
  - The theoretical exercises have to be presented in front of the class.
- Programming:
  - 60% of all programming assignments have to be presented during the tutorial.

If someone receives a mark between 1.3 and 4.0 in the final exam, the mark will be
improved by 0.3 and 0.4, respectively. (Note that marks of 1.0 and 5.0 cannot
be improved.)
The course is intended for Master students.

Prerequisites:
- Basic Mathematics:
  - multivariate analysis and
  - linear algebra.
- Basic Computer Science:
  - programming skills (Matlab, C/C++) and
  - basic knowledge in algorithms and data structures (e.g., dynamic programming).

Object detection

Digital image

A color digital image with red, green and blue (RGB) channels can be defined as a function

\[ I : \Omega \subset \mathbb{Z}^2 \to [0, 255]^3 \times N \subset \mathbb{Z}^{3 \times N} , \ N = |\Omega| . \]

For an image with size of \( W \times H \), the domain

\[ \Omega = \{1, \ldots, W\} \times \{1, \ldots, H\} . \]

An image \( I \) is composed of pixels (picture element or picture cell)

\[ i \in \Omega , \quad I(i) = (r_i, g_i, b_i) \in [0, 255]^3 . \]

For more details you may refer to the course on Computer Vision I: Variational Methods

Object detection

We address the problem of binary image segmentation, where we also assume that we are provided with non-local parameters that are known a priori.

For example, one can assume prior knowledge about the shape of the foreground.

Exemplar binary segmentation of cars assuming shape prior

The goal is to give a label $y_i \in \mathcal{L} = \{1, 2, \ldots, r\}$ for each pixel $i$ according to its semantic meanings. The labeling is defined as $L : \Omega \rightarrow \mathcal{L}^N$.

Exemplar semantic segmentations

Online demo: http://www.robots.ox.ac.uk/~szheng/crfasrnndemo

Suppose that we are given two cameras looking at parallel direction. Let $C_{\text{left}}$ be the origin of the coordinate system and assume that the image planes are co-planar and parallel to the $x$ and $y$ axis.

The intersection of the triangle $\triangle(C_{\text{left}}, P, C_{\text{right}})$ and the plane including the images planes is the segment $P_{P_{\text{left}}}$. Therefore $P_{P_{\text{left}}}$ is parallel to the $x$-axis.

For more details you may refer to the course on Computer Vision II: Multiple View Geometry.

The goal is to recognize an articulated object (e.g., human body) with different connecting parts (e.g., head, torso, left arm, right arm, left leg, right leg).

An object is composed of a number of rigid parts, where each part is modeled as a rectangle.

Each part is modeled as a rectangle parameterized by $(x, y, s, \theta)$, where
- $(x, y)$ means the center of the rectangle,
- $s \in [0, 1]$ is a scaling factor, and
- the orientation is given by $\theta$.

In overall, we have a four-dimensional pose space as label space $\mathcal{L}$.

We denote the label (i.e. locations) of a part by $l_i = (x_i, y_i, s_i, \theta_i) \in \mathcal{L}$.

An object (e.g., human body) is given by a configuration

$$I = (l_1, \ldots, l_N),$$

where $l_i$ specifies the location of part $v_i$ and $N$ is the number of object parts.
An experiment is a (random) process that can be infinitely many times repeated and has a well-defined set of possible outcomes. In case of repeated experiments the individual repetitions are also called trials.

Example: throwing two “fair dice” (i.e. we assume equally likely chance of landing on any face) with six faces.

The sample space, denoted by $\Omega$, is the set of possible outcomes.

Example: $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$.

A set of outcomes $A \subseteq \Omega$ is called an event. An atomic event is an event that contains a single outcome $\omega \in \Omega$.

Example: $A = \{(i, j) : i + j = 11\}$, i.e. the sum of the numbers showing on the top is equal to eleven.

\[ \begin{align*}
\Omega &= \{(i, j) : 1 \leq i, j \leq 6\} \\
A &= \{(i, j) : i + j = 11\}
\end{align*} \]

The $\emptyset$ is called the impossible event; and $\Omega$ is the sure event.

Let $A$ and $B$ be two events from an sample space $\Omega$. We will use the following notations:

- $A$ does not occur: $\bar{A} = \Omega \setminus A$
- either $A$ or $B$ occur: $A \cup B$
- both $A$ and $B$ occur: $A \cap B$
- $A$ occurs and $B$ does not: $A \setminus B$

A probability space represents our uncertainty regarding an event. A triple $(\Omega, P, \mathcal{F})$ is called a discrete probability space, if

- $\Omega$ is not empty and countable (i.e. $|\Omega| = |\mathbb{N}|$), and
- $A$ is the power set $\mathcal{P}(\Omega)$ (i.e. the set of all subsets of $\Omega$), and
- $P : \mathcal{F} \rightarrow \mathbb{R}$ is a function, called a probability measure, with the following properties:
  1. $P(A) \geq 0$ for all $A \in \mathcal{F}$
  2. $P(\Omega) = 1$
  3. $\sigma$-additivity holds: if $A_1, \ldots, A_n \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$,
     then $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

The conditions 1-3. are called the Kolmogorov’s axioms.

For this case a discrete probability space $(\Omega, P(\Omega), P)$ is given by

- Sample space: $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$, $|\Omega| = 36$.
- $P(\Omega) = \{(1,1), \ldots, (1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,6)\}$.
- The (uniform) probability measure for all $A \in P(\Omega)$
  \[ P(A) = \frac{|A|}{36} , \]

where $k$ is the number of atomic events in $A$. Such probability space (i.e. all outcomes are equally likely) is called the classical probability space.

Example: Let $A$ denote the event that “the sum of the numbers showing on the top is equal to eleven”, that is $A = \{(i, j) : i + j = 11\} = \{(5, 6), (6, 5)\}$.

Hence, $P(A) = P((5, 6), (6, 5)) = \frac{2}{36}$.

Some simple consequences of the axioms

The following rules are frequently used in applications:

- $P(A) = 1 - P(\bar{A}) = 1 - P(\Omega, A)$.
  
  \textbf{Proof.} Note that $A$ and $\Omega \setminus A$ are disjoint.
  
  \[ 1 = P(\Omega) = P(A \cup (\Omega, A)) = P(A) + P(\Omega, A) . \]

- $P(\emptyset) = 0$.
  
  \textbf{Proof.} $P(\emptyset) = 1 - P(\Omega, \emptyset) = 1 - 1 = 0$.

- If $A \subseteq B$, then $P(A) \leq P(B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(A \cup B) \leq P(A) + P(B)$.
- $\cdots$

(See the proofs in the Exercise.)

Conditional probability allows us to reason with partial information. If $P(B) > 0$, the conditional probability of $A$ given $B$ is defined as

\[ P(A \mid B) = \frac{P(A \cap B)}{P(B)} . \]

This is the probability that $A$ occurs, given we have observed $B$, i.e. we know the experiment’s actual outcome will be in $B$.

Note that the axioms and rules of probability theory are fulfilled for the conditional probability. (e.g., $P(A \mid B) = 1 - P(\bar{A} \mid B)$.)
Consider the problem of binary segmentation, that is each pixel belongs to either the foreground or the background.

Let us define a pixel to be “bright”, if all its (RGB) intensities are at least 128, otherwise the given pixel is considered to be “dark”.

Assume we are given the following table with probabilities:

<table>
<thead>
<tr>
<th></th>
<th>Bright</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreground</td>
<td>0.163</td>
<td>0.006</td>
</tr>
<tr>
<td>Background</td>
<td>0.116</td>
<td>0.715</td>
</tr>
</tbody>
</table>

**Question:** What is the probability of a pixel belongs to the foreground, when it is dark?

Let A denote the event that “the pixel belongs to the foreground” and let B denote the event that “dark”. Then

\[ P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{0.161}{0.279} \approx 0.584. \]

### Independence

Two events A and B are independent, denoted by \( A \perp B \), if

\[ P(A | B) = P(A), \]

equivalently,

\[ P(A \cap B) = P(A)P(B). \]

If A and B are independent, learning that \( B \) happened does not make \( A \) more or less likely to occur.

**Example:** Suppose we roll a “fair” die. Let us consider the events A denoting “the die outcome is even” and \( B \) denoting “the die outcome is either 1 or 2”.

\[ P(A) = \frac{1}{2} \] and \( P(B) = \frac{1}{6} \). Moreover \( A \cap B \) means the event that the outcome is two, that is \( P(A \cap B) = \frac{1}{6} \).

\[ P(A | B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{6} = P(A)P(B) \quad \Rightarrow \quad A \text{ and } B \text{ are independent.} \]

### Conditional independence

Let A, B and C be events. A and B are conditionally independent given \( C \), denoted by \( A \perp B \mid C \), if

\[ P(A | B \cap C) = P(A | C), \]

equivalently,

\[ P(A \cap B | C) = P(A | C)P(B | C). \]

A and \( B \) are conditionally independent given \( C \) means that once we learned \( C \), learning \( B \) gives us no additional information about \( A \).

**Example:** The operation of a car’s starter motor \( A \) is conditionally independent its radio \( B \) given the status of the battery \( C \).

### Summary

- A discrete probability space is a triple \((\Omega, P(\Omega), P)\), where \( \Omega \) is countable, \( P: P(\Omega) \to \mathbb{R} \) is \( \sigma \)-additive and for all \( A \in P(\Omega) \) and \( 0 \leq P(A) \leq 1 \leq P(\Omega) \).
- The conditional probability of \( A \) given \( B \), is the probability that \( A \) occurs, given we have observed \( B \).
- If A and B are independent \( (A \perp B) \), learning that \( B \) happened does not make \( A \) more or less likely to occur.
- A and B are conditionally independent given \( C \) \((A \perp B \mid C)\) means that once we learned \( C \), learning \( B \) gives us no additional information about \( A \).

In the next lecture we will learn about

- Random variables
- Probability distributions
- Probabilistic Graphical Models

### A brain teaser

Suppose you are on a game show and you are given the choice of three doors: Behind one door is a car; behind the others, goats.

You pick a door, say No. 1, and the host, who knows what is behind the doors, opens another door, say No. 3, which has a goat.

Then he says to you, "Do you want to pick door No. 2?"

**Question:** Is it to your advantage to switch your choice?