Lab 1: Inference and Representation - Concept Check

Vlad Kobzar

NYU CDS

September 11, 2017

Welcome!

- Lab Instructor: Vlad Kobzar (vk283@nyu.edu)
- Office: 60 5th Ave, Rm 737

Today we will review the following:

1. Conditional probability, LOTP and the Bayes rules
2. Conditional independence
3. Graph representations of conditional probability distributions

Definition

Given a joint distribution of X and Y, what if we want to know the distribution of X when Y is set to a particular value? More precisely:

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

Note that this is a univariate distribution, defined for each value of y

Basic probabilistic model: Flights and rain

Probabilistic model for late arrivals at an airport

$$\Omega = \{\text{late and rain}, \text{late and no rain}, \\ \text{on time and rain}, \text{on time and no rain}\}$$

$$\mathrm{P}(\text{late}, \text{no rain}) = \frac{2}{20}, \quad \mathrm{P}(\text{on time}, \text{no rain}) = \frac{14}{20},$$

$$\mathrm{P}(\text{late}, \text{rain}) = \frac{3}{20}, \quad \mathrm{P}(\text{on time}, \text{rain}) = \frac{1}{20}$$

$$\mathrm{P}(\text{late}|\text{rain}) \ ?$$

Basic probabilistic model: Flights and rain

Probabilistic model for late arrivals at an airport

$$\mathrm{P}(\text{late}|\text{rain}) = \frac{\mathrm{P}(\text{late}, \text{rain})}{\mathrm{P}(\text{rain})} = \frac{3/20}{3/20 + 1/20} = \frac{3}{4}$$

and similarly $\mathrm{P}(\text{late}|\text{no rain}) = 1/8$.

LOTP

Chain Rule:

$$p(x_1, \ldots, x_n) = \underbrace{\underbrace{p(x_1)p(x_2|x_1)}_{p(x_1,x_2)} p(x_3|x_1, x_2)}_{p(x_1,x_2,x_3)\ldots} \ldots p(x_n|x_1, x_2, \ldots, x_{n-1})$$

Specifies how you can write the joint probability as the product of conditional probabilities.

LOTP

Sometimes, estimating the probability of a certain event directly may be more challenging than estimating its probability conditioned on other simpler events that cover the whole sampling space. The Law of Total Probability, which follows from the Chain Rule allows us to pool conditional probabilities together, weighting them by the probability of the individual events in the partition, to compute the probability of the event in interest. If $A_1, A_2, \ldots \in \mathcal{F}$ is a partition of Ω, for any set $S \in \mathcal{F}$

$$\mathrm{P}(S) = \sum_i \mathrm{P}(S \cap A_i) = \sum_i \mathrm{P}(A_i)\,\mathrm{P}(S|A_i)$$

LOTP: Flights and rain (continued)

$$\mathrm{P}(\text{rain}) = 0.2$$
$$\mathrm{P}(\text{late}|\text{rain}) = 0.75$$
$$\mathrm{P}(\text{late}|\text{no rain}) = 0.125$$

$$\mathrm{P}(\text{late}) \ ?$$

LOTP: Flights and rain (continued)

The events rain and no rain are disjoint and cover the whole sample space, so they form a partition. We can consequently apply the Law of Total Probability to determine

$$\begin{aligned}\mathrm{P}(\text{late}) &= \mathrm{P}(\text{late}|\text{rain})\,\mathrm{P}(\text{rain}) + \mathrm{P}(\text{late}|\text{no rain})\,\mathrm{P}(\text{no rain}) \\ &= 0.75 \cdot 0.2 + 0.125 \cdot 0.8 = 0.25.\end{aligned}$$

Bayes' Rule

$p(A|B) \neq p(A|B)$, however, it is possible to *invert* conditional probabilities by the Bayes' Rule.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

- Useful when $p(y|x)$ is more natural to measure or estimate than $p(x|y)$
- Proof: From chain rule and definition of conditional distribution

Example: Flights and rain (continued)

$$P(\text{rain}) = 0.2$$
$$P(\text{late}|\text{rain}) = 0.75$$
$$P(\text{late}|\text{no rain}) = 0.125$$

$$P(\text{rain}|\text{late}) \; ?$$

Example: Flights and rain (continued)

$$
\begin{aligned}
\mathrm{P}(\text{rain}|\text{late}) &= \frac{\mathrm{P}(\text{rain}, \text{late})}{\mathrm{P}(\text{late})} \\
&= \frac{\mathrm{P}(\text{late}|\text{rain})\,\mathrm{P}(\text{rain})}{\mathrm{P}(\text{late}|\text{rain})\,\mathrm{P}(\text{rain}) + \mathrm{P}(\text{late}|\text{no rain})\,\mathrm{P}(\text{no rain})} \\
&= \frac{0.75 \cdot 0.2}{0.75 \cdot 0.2 + 0.125 \cdot 0.8} = 0.6
\end{aligned}
$$

Basic idea

- We will represent the world by a collection of jointly distributed random variables $(X_1, ..., X_n)$, which can be estimated by an exponential amount of data 2^n (if $|Val(X_i)| = 2$).
- If $X_1 ... X_n$ are conditionally independent given Y, the distribution can be estimated by an amount of data linear in n and has predictive power.

$$p(x_1, x_2, ..., x_n, y) = p(y) \prod_{i=1}^{n} p(x_i|y)$$

Does conditional independence imply independence?

Probabilistic model for taxi availability, flight delay and weather:

$$\mathrm{P}(\text{rain}) = 0.2$$
$$\mathrm{P}(\text{late}|\text{rain}) = 0.75$$
$$\mathrm{P}(\text{late}|\text{no rain}) = 0.125$$
$$\mathrm{P}(\text{taxi}|\text{rain}) = 0.1$$
$$\mathrm{P}(\text{taxi}|\text{no rain}) = 0.6$$

Given *rain* and *no rain*, *late* and *taxi* are conditionally independent. Are they also independent?
$\mathrm{P}(\text{taxi}) = \mathrm{P}(\text{taxi}|\text{late})$?

Conditional independence does not imply independence

$$
\begin{aligned}
\mathrm{P}(\text{taxi}) &= \mathrm{P}(\text{taxi}|\text{rain})\,\mathrm{P}(\text{rain}) + \mathrm{P}(\text{taxi}|\text{no rain})\,\mathrm{P}(\text{no rain}) \\
&= 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \\
\mathrm{P}(\text{taxi}|\text{late}) &= \frac{\mathrm{P}(\text{taxi}, \text{late})}{\mathrm{P}(\text{late})} \\
&= \frac{\mathrm{P}(\text{taxi}, \text{late}, \text{rain}) + \mathrm{P}(\text{taxi}, \text{late}, \text{no rain})}{\mathrm{P}(\text{late})} \\
&= \frac{\mathrm{P}(\text{t}|\text{l}, \text{r})\,\mathrm{P}(\text{l}|\text{r})\,\mathrm{P}(\text{r}) + \mathrm{P}(\text{t}|\text{l}, \text{no r})\,\mathrm{P}(\text{l}|\text{no r})\,\mathrm{P}(\text{no r})}{\mathrm{P}(\text{l})} \\
&= \frac{\mathrm{P}(\text{taxi}|\text{r})\,\mathrm{P}(\text{late}|\text{r})\,\mathrm{P}(\text{r}) + \mathrm{P}(\text{taxi}|\text{no r})\,\mathrm{P}(\text{late}|\text{no r})\,\mathrm{P}(\text{no r})}{\mathrm{P}(\text{late})} \\
&= \frac{0.1 \cdot 0.75 \cdot 0.2 + 0.6 \cdot 0.125 \cdot 0.8}{0.25} = 0.3
\end{aligned}
$$

Does independence imply conditional independence?

Probabilistic model for mechanical problems, weather and delays

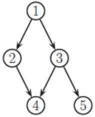
$$
\begin{aligned}
\mathrm{P}(\text{rain}) &= 0.2 \\
\mathrm{P}(\text{late}|\text{rain}) &= 0.75 \\
\mathrm{P}(\text{late}|\text{no rain}) &= 0.125 \\
\mathrm{P}(\text{problem}) &= 0.1 \\
\mathrm{P}(\text{late}|\text{problem}) &= 0.7 \\
\mathrm{P}(\text{late}|\text{no problem}) &= 0.2 \\
\mathrm{P}(\text{late}|\text{no rain}, \text{problem}) &= 0.5
\end{aligned}
$$

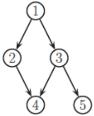
problem and *no rain* are independent. Are they also conditionally independent given *late*?
$\mathrm{P}(\text{problem}|\text{late}, \text{no rain}) = \mathrm{P}(\text{problem}|\text{late})$?

Independence does not imply conditional independence

$$
\begin{aligned}
\mathrm{P}(\text{problem}|\text{late}) &= \frac{\mathrm{P}(\text{late}, \text{problem})}{\mathrm{P}(\text{late})} \\
&= \frac{\mathrm{P}(\text{late}|\text{p})\,\mathrm{P}(\text{p})}{\mathrm{P}(\text{late}|\text{p})\,\mathrm{P}(\text{p}) + \mathrm{P}(\text{late}|\text{no p})\,\mathrm{P}(\text{no p})} \\
&= \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28 \\
\mathrm{P}(\text{problem}|\text{late}, \text{no rain}) &= \frac{\mathrm{P}(\text{late}, \text{no rain}, \text{problem})}{\mathrm{P}(\text{late}, \text{no rain})} \\
&= \frac{\mathrm{P}(\text{late}|\text{no rain}, \text{problem})\,\mathrm{P}(\text{no rain})\,\mathrm{P}(\text{problem})}{\mathrm{P}(\text{late}|\text{no rain})\,\mathrm{P}(\text{no rain})} \\
&= \frac{0.5 \cdot 0.1}{0.125} = 0.4
\end{aligned}
$$

We represent probability distributions by directed acyclic graphs (Bayesian networks) (and we may encounter undirected graphs (Markov random fields) later in the course)
Write down the joint conditionally independent probability distribution encoded by the DAG below.





The joint distribution is given by

$$\begin{aligned} p(x_{1:5}) &= p(x_1)p(x_2|x_1)p(x_3|x_1,x_2)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_2,x_3,x_4) \\ &= p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2,x_3)p(x_5|x_3) \end{aligned}$$

References/Acknowledgments

- Sections 1.2 and 1.3 in: http://www.cims.nyu.edu/~cfgranda/pages/stuff/probability_stats_for_DS.pdf
- Carlos Ferndandez-Granda, Rahul Krishnan, Rachel Hodos, whose slides I used for this presentation
- Murphy, 10.1.5 (in re DAGs)