Machine Learning for Neuroscience

Slides and notebooks: `https://github.com/PBarnaghi/ML4NS`
Notes by: *Nan Fletcher-Lloyd, Payam Barnaghi*

Lecture 3. Probability and Information Theory

In lecture 3, we cover some of the basic concepts of probability and information theory.

With sufficient data, statistics can enable us to calculate probabilities using real-world observations. Here, probability provides the theory while statistics provides the tools to test that theory using data. In the end, descriptive statistics such as the mean and standard deviation of the data become proxies for theoretical. This is because real-world probabilities are often quite difficult to calculate. As such, we rely on statistics and data. With more and more data, we can become more confident that what we create models that represent the true probability of these events occurring.

In this lecture, we first introduce probability with regards to how likely it is that a discrete random variable (or continuous) is to take on each of its possible states, otherwise known as the probability distribution or the probability density function. We further introduce the cumulative distribution function for continuous random variables. We also provide a description of joint probabilities and conditional probability. We further expand on the two key characteristics of a probability distribution (mean and standard deviation). Finally, we consider the different types (with examples) of probability distributions, including the Bernoulli, Binomial, Multinomial and Multinoulli, and Gaussian (normal) and standard normal distributions.

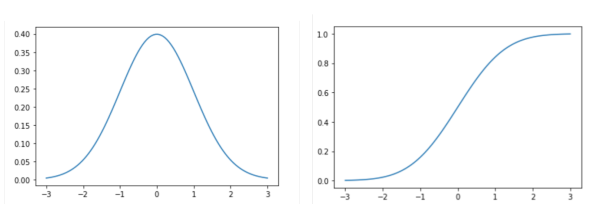

Figure 3.1. Probability Density Function vs Cumulative Density Function

Figure 3.1 illustrates the difference between a probability density function (in the case of discrete random variables) and a cumulative density function (in the case of continuous random variables).

This lecture also introduces information theory, with the basic intuition that learning that an unlikely event has occurred is more informative than learning that a likely event has occurred. Information theory has three main properties: likely events should have low information content, and in the extreme case, events that are guaranteed to happen should have no information content whatsoever, less likely events should have higher information content, and independent events should have additive information. Finally, we introduce conditional independence in the context of using Markov Chains (describing the possible sequences of events) to model behaviour. Figure 3.2. illustrates the Markov chain representation of in-home movements in a household of a person living with dementia.

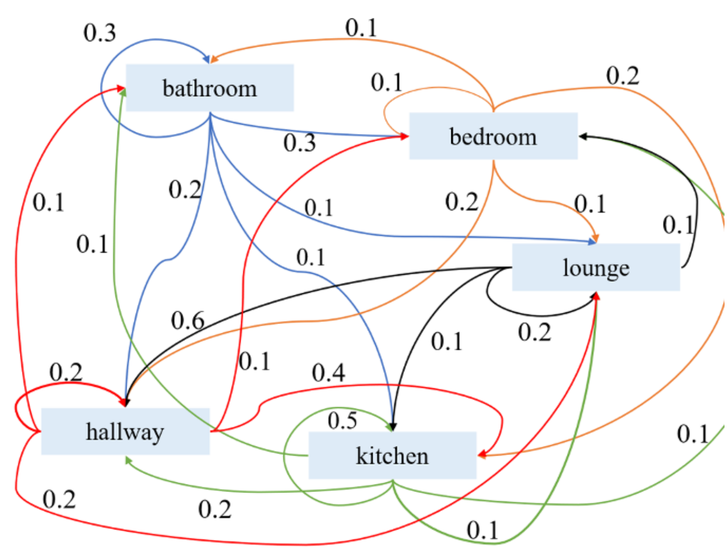

Figure 3.2. Figure 2. A Markov chain representation of in-home movements in a household of a person living with dementia.

The tutorial and assessment are combined with the next lecture's tutorial and assessment and aim to build your understanding of how we can use statistics as a tool to calculate probabilities using real-world observations.