Bayesian Networks

Adam Massmann

Water Center NN Meetings

Week 5 - Oct. 17th

Introduction - Information Theory

- Consider the ammount of information gained/learned by an event or observation (we'll call x). We would receive more new information from a very suprising event than an event we expect.
- So say we want to quantify the "ammount of information" contained in an event. This should then be a function of the probability of the event ($p(x)$). So the ammount of information we'll call a function $h(\cdot)$ will be a function of $p(x)$.

Guidance for the functional from of $h(\cdot)$

- If two events x and y are independent, then the ammount of information gained by both events should be $h(x, y) = h(x) + h(y)$.
- We also know that the joint probability of x and y's occurence would be: $p(x, y) = p(x)\ p(y)$.
- So the question is, what function h satisfies: $h(p(x, y)) = h(p(x)\ p(y)) = h(p(x)) + h(p(y))$?

Information Entropy

- $h(\cdot) = \log(p(\cdot))$ satisfies $h(p(x, y)) = h(p(x)\, p(y)) = h(p(x)) + h(p(y))$
- Its desirable for h to be postiive, so because $0 \leq p \leq 1$, lets make it $h(\cdot) = -\log p(\cdot)$.
- Now say we have a bunch of random variables x for which we want to know the average ammount of information (i.e. expectation of $h(x)$). This would be given by:

$$H[x] = -\sum_{x} p(x) \log p(x)$$

- This is known as the *entropy* of x.
- Extending this to continuous variabiles gives the *differential entropy*:

$$H[x] = -\int p(x)\, \log p(x)\, dx$$

Tables and Figures

- Use `tabular` for basic tables — see Table 1, for example.
- You can upload a figure (JPEG, PNG or PDF) using the files menu.
- To include it in your document, use the `includegraphics` command (see the comment below in the source code).

Item	Quantity
Widgets	42
Gadgets	13

Table 1: An example table.

Readable Mathematics

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i}^{n} X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.