CIS 291d. Assignment 1

Out: *Tue Jan 09*
Due: *Fri Jan 19*

1.1 Entropy

Let X be a discrete random variable with $P(X=x_i) = p_i$ for $i \in \{1, 2, \ldots, n\}$. The entropy $\mathcal{H}[X]$ of the random variable X is a measure of its uncertainty. It is defined as:

$$\mathcal{H}[X] \;=\; -\sum_{i=1}^{n} p_i \log p_i.$$

Show that the entropy $\mathcal{H}[X]$ is maximized when $p_i = \frac{1}{n}$ for all i. You should do this by computing the gradient with respect to p_i and using Lagrange multipliers to enforce the constraint that $\sum_i p_i = 1$. Later in the course, we will use similar calculations for learning probabilistic models.

1.2 Kullback-Leibler distance

Consider two discrete probability distributions, p_i and q_i, with $\sum p_i = \sum q_i = 1$. The Kullback-Leibler (KL) distance between these distributions (also known as the relative entropy) is defined as:

$$\mathrm{KL}(p, q) = \sum_i p_i \log(p_i/q_i).$$

(a) By sketching graphs of $\log(x)$ and $x-1$, verify the inequality:

$$\log(x) \;\leq\; x-1,$$

with equality if and only if $x=1$. Confirm this result by differentiation of $\log(x)-(x-1)$.

(b) Use the previous result to prove that:

$$\mathrm{KL}(p, q) \;\geq\; 0,$$

with equality if and only if the two distributions p_i and q_i are equal.

(c) Using the inequality in (a), as well as the simple equality $\log x = 2\log\sqrt{x}$, derive the tighter lower bound:

$$\mathrm{KL}(p, q) \;\geq\; \sum_i (\sqrt{p_i} - \sqrt{q_i})^2.$$

(d) Provide a counterexample to show that the KL distance is not a symmetric function of its arguments:

$$\mathrm{KL}(p, q) \;\neq\; \mathrm{KL}(q, p).$$

Despite this asymmetry, it is still common to refer to $\mathrm{KL}(p, q)$ as a measure of distance. Many algorithms in machine learning are based on minimizing KL distances between probability distributions.

1.3 Conditioning on background evidence [RN 13.9]

It is often useful to consider the impact of specific events in the context of general background evidence, rather than in the absence of information.

(a) Denoting such evidence by E, prove the conditionalized version of the product rule:

$$P(X,Y|E) \;=\; P(X|Y,E)P(Y|E).$$

(b) Also, prove the conditionalized version of Bayes rule:

$$P(X|Y,E) \;=\; \frac{P(Y|X,E)P(X|E)}{P(Y|E)}.$$

1.4 Conditional independence [RN 13.10]

Show that the following three statements about random variables X, Y, and Z are equivalent:

$$\begin{aligned} P(X,Y|Z) &= P(X|Z)P(Y|Z) \\ P(X|Y,Z) &= P(X|Z) \\ P(Y|X,Z) &= P(Y|Z) \end{aligned}$$

You should become fluent with all these ways of expressing that X is conditionally independent of Y given Z.

1.5 Creative writing

Attach events to the binary random variables X, Y, and Z that are consistent with the following patterns of commonsense reasoning. You may use different events for the different parts of the problem.

(a) Cumulative evidence:

$$P(X\!=\!1) < P(X\!=\!1|Y\!=\!1) < P(X\!=\!1|Y\!=\!1,Z\!=\!1)$$

(b) Explaining away:

$$\begin{aligned} P(X\!=\!1|Y\!=\!1) &> P(X\!=\!1), \\ P(X\!=\!1|Y\!=\!1,Z\!=\!1) &< P(X\!=\!1|Y\!=\!1) \end{aligned}$$

(c) Conditional independence:

$$\begin{aligned} P(X\!=\!1,Y\!=\!1) &\neq P(X\!=\!1)P(Y\!=\!1) \\ P(X\!=\!1,Y\!=\!1|Z\!=\!1) &= P(X\!=\!1|Z\!=\!1)P(Y\!=\!1|Z\!=\!1) \end{aligned}$$