Assignment Four

Matthew C. Scicluna
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal
Montréal, QC H3T 1J4
matthew.scicluna@umontreal.ca

November 5, 2017

1 Entropy and Mutual Information

Let X be a discrete random variable on a finite space $\mathcal{X}$ with $|\mathcal{X}| = k$.

1. (a) Prove that the entropy $H(X) \geq 0$, with equality only when X is a constant.

   PROOF: WLOG we can assume that $p(x) > 0 \ \forall x \in \mathcal{X}$, (using that $0 \cdot log 0 = 0$). We have that $H(X) = -\sum_x p(x) log p(x) = \sum_x p(x) log p(x)^{-1} \geq 0$, since $p(x) > 0$ and $p(x)^{-1} \geq 1$. If $H(X) = 0$ then $\exists \alpha$ such that $p(\alpha)^{-1} = 1 \Rightarrow p(\alpha) = 1$. Hence X must be a constant, as needed.

   (b) Let $X \sim p$ and q be the Uniform distribution on $\mathcal{X}$. What is the relation between $D(p||q)$ and $H(X)$.

   CLAIM: $D(p||q) = -H(X) + logk$

   PROOF:

$$\begin{aligned} D(p||q) &= \sum_x p(x) log \frac{p(x)}{q(x)} \\ &= \sum_x p(x) log p(x) - \sum_x p(x) log q(x) \\ &= -H(X) + \sum_x p(x) log k \\ &= -H(x) + log k \end{aligned}$$

(c) An upper bound for $H(X)$ is $logk$ since $H(X) = logk - D(p||q) \Rightarrow H(X) \leq logk$.

We consider a pair of discrete random variables (X_1, X_2) defined over the finite set $\mathcal{X}_1 \times \mathcal{X}_2$. Let $p_{1,2}$, p_1 and p_2 denote respectively the joint distribution, the marginal distribution of X_1 and the marginal distribution of X_2. Define the mutual information as:

$$I(X_1, X_2) = \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1, x_2) log \frac{p_{1,2}(x_1, x_2)}{p_1(x_1)p_2(x_2)}$$

We again assume WLOG that $p(x_1, x_2) > 0 \ \forall (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$

2. (a) CLAIM: $I(X_1, X_2) \geq 0$

   PROOF: Notice that $I(X_1, X_2) = D(p_{1,2}||p_1 p_2) \geq 0$ by the positiveness of $D(\cdot||\cdot)$.

   (b) We want to express $I(X_1, X_2)$ as a function of $H(X_1), H(X_2)$ and $H(X_1, X_2)$.

$$\begin{aligned}
I(X_1, X_2) &= \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1, x_2) log \frac{p_{1,2}(x_1, x_2)}{p_1(x_1)p_2(x_2)} \\
&= \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1, x_2) log p_{1,2}(x_1, x_2) - p_{1,2}(x_1, x_2) log p_1(x_1)p_2(x_2) \\
&= -H(X_1, X_2) - \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} \Big(p_1(x_1)p_2(x_2) log p_1(x_1) - p_1(x_1)p_2(x_2) log p_2(x_2) \Big) \\
&= -H(X_1, X_2) - \sum_{j=1}^{2} \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_1(x_1)p_2(x_2) log p_j(x_j) \\
&= -H(X_1, X_2) - \sum_{j=1}^{2} \sum_{x_j\in\mathcal{X}_j} p_j(x_j) log p_j(x_j) \\
&= -H(X_1, X_2) + H(X_1) + H(X_2)
\end{aligned}$$

   And so we can represent $I(X_1, X_2)$ using $H(X_1), H(X_2)$ and $H(X_1, X_2)$, as needed.

   (c) From the previous result we have that $I(X_1, X_2) \geq 0 \Rightarrow H(X_1) + H(X_2) \geq H(X_1, X_2)$, and so the maximal entropy of (X_1, X_2) is $H(X_1) + H(X_2)$. By definition this only occurs when $I(X_1, X_2) = 0$, which only occurs if $p_{1,2}(x_1, x_2) = p_1(x_1)p_2(x_2) \ \forall x_1, x_2 \in \mathcal{X}_1 \times \mathcal{X}_2$. This can be seen directly from the definition of I and using the strict positivity of $p(x_1, x_2)$.

2 HMM – Implementation

We use an HMM model to account for the possible temporal structure of some data. We consider the following HMM model: the chain $(z_t)_{t=1}^T$ has $K = 4$ possible states, with an initial probability distribution $\pi \in \Delta_4$ and a probability transition matrix $A \in \mathbb{R}^{4\times 4}$ where $A_{ij} = p(z_t = i|z_{t-1} = j)$ and conditionally on the current state z_t, we have observations obtained from Gaussian emission probabilities $x_t|(z_t = k) \sim N(x_t|\mu_k, \Sigma_k)$.