Homework II

Due: Jan. 30, 2007 (midnight)

January 25, 2008

Problem 1 (20pt)

• Show $H(X, Y) \leq H(X) + H(Y)$
• Show $I(X; Y) = H(X) + H(Y) - H(X, Y)$
• Show $H(p) = \sum_{i=1}^{m} p_i \log(1/p_i)$ is a concave function for variables $p_1, \ldots, p_m$ within the domain $\{p = (p_1, \ldots, p_m) | \sum_{i=1}^{m} p_i = 1, p_i \geq 0, i = 1, \ldots, m\}$.

Hint: You need to prove for any two distributions $p = (p_1, \ldots, p_m)$ and $q = (q_1, \ldots, q_m)$, you have

$$\theta H(p) + (1 - \theta) H(q) \leq H(\theta p + (1 - \theta)q)$$

holds for any $0 \leq \theta \leq 1$. The key to the proof is to utilize the inequality

$$\sum_{i=1}^{m} p_i \log(1/p_i) \leq \sum_{i=1}^{m} p_i \log(1/q_i)$$

Problem 2: (20pt)

• Compute the Bregman distance $B_{\varphi}(p||q)$ between distribution $q = (q_1, q_2, \ldots, q_m)$ and $p = (p_1, p_2, \ldots, p_m)$ for function (1) $\varphi(q) = \|q\|_2^2$, and (2) $\varphi(q) = -H(q)$ where $H(q)$ is the entropy function.
• Show $B_{\varphi}(x||y) = B_{\varphi}(x||z) + B_{\varphi}(z||y) - (x - z)^T (\nabla \varphi(y) - \nabla \varphi(z))$ where $B_{\varphi}(-||\cdot)$ is a Bregmen distance and $\varphi(\cdot)$ is a convex function.
• Prove the generalized Pythagoras Theorem, i.e,

$$B_{\varphi}(x||y) \geq B_{\varphi}(x||P_{\Omega}(y)) + B_{\varphi}(P_{\Omega}(y)||y)$$

for any point $x \in \Omega$ where $\Omega$ is a convex region. $P_{\Omega}(y)$ denotes the projection of $y$ onto the convex region $\Omega$ and is defined as

$$P_{\Omega}(y) = \arg \min_{x \in \Omega} B_{\varphi}(y||x)$$

Figure 1 illustrates the basic idea of the theorem.

Problem 3 (20pt)

Consider that somebody has both a die and a coin. Each time, he can choose to roll the die or to flip the coin. The decision is based on a Bernoulli distribution. The observed number for flipping the coin is either 0 or 1, and the observed number for rolling the die is either 0, 1, 2, 3, 4, or 5. By repeating this experiment for $m$ times, you will observed a sequence of numbers $(n_1, n_2, \ldots, n_m)$. The observation data can be found in http://www.cse.msu.edu/~cse847/assignments/mixture_data.txt.
Assume that each experiment is independent from the previous one, and meanwhile both the coin and the die are fair with equal probability for each number. Compute the probability $p$ in the Bernoulli distribution using maximum likelihood estimation that fits the observations.

Still assume that each experiment is independent from others. But assume that we only know that the die is fair one but not the coin. Re-compute your answer for the probability $p$ in Bernoulli distribution, and in the meantime estimate the bias of the coin $b$, which is the probability for observing 1 when randomly flipping the coin.