(520|600).666
Information Extraction from Speech and Text

Homework # 8

Due April 9, 2009.

Review Chapter 2 from *Elements of Information Theory* by Tom Cover and Joy Thomas.

1. **Entropy of discrete random variables**:

   (a) Starting with the definition

   $$H(X_1, X_2, \ldots, X_n) = \sum_{x_1, x_2, \ldots, x_n \in \mathcal{X}^n} p(x_1, x_2, \ldots, x_n) \log \frac{1}{p(x_1, x_2, \ldots, x_n)},$$

   prove the chain rule of entropy:

   $$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_1, X_2, \ldots, X_{n-1}).$$

   Try $n = 2$ first to see the basic application of Bayes' rule.

   (b) Given two random variables X and Y, draw a Venn diagram to describe the relationships between $H(X,Y)$, $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$ and $I(X;Y)$.

   (c) Establish the relationship between the entropy $H(\mathbf{p})$ of a probability mass function $\mathbf{p}$ and the Kullback-Leibler distance $D(\mathbf{p}\|\mathbf{u})$ between $\mathbf{p}$ and the uniform distribution $\mathbf{u}$.

2. **Conditional Mutual Information**: Define the *conditional* mutual information between two random variables X and Y, given a third random variable Z, as

   $$I(X;Y|Z) = \sum_{z \in \mathcal{Z}} p(z) I(X;Y|Z=z) = \sum_{z \in \mathcal{Z}} p(z) \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)}.$$

   (a) Manipulate the expression above to show that

   $$I(X;Y|Z) = H(X|Z) - H(X|Y,Z).$$

   (b) From the expression above, derive the chain-rule of mutual information

   $$I(X_1, X_2, \ldots, X_n; Y) = I(X_1;Y) + I(X_2;Y|X_1) + \ldots + I(X_n;Y|X_1, \ldots, X_{n-1}).$$

These are standard results whose proofs are easily found elsewhere. To gain understanding of the concepts, it is important that you derive these results yourself, not copy the derivation from a textbook. Please make the effort.