Homework 1

10-704 Information Processing and Learning

Instructor: Aarti Singh

The HW is worth 40 pts and is **due on Feb 3 at noon**. Hand in to: Michelle Martin GHC 8001. If she is not around, note down the time on your HW sheet and slide it under her door.

1. [10 pts] *Information content*
   You are given 12 balls, all equal in weight except for one that is either heavier or lighter. You are also given a two-pan balance to use. In each use of the balance you may put any number of the 12 balls on the left pan, and the same number on the right pan. There are three possible outcomes: either the weights are equal, or the balls on the left are heavier, or the balls on the left are lighter. The goal is to come up with a strategy to determine which is the odd ball and whether it is heavier or lighter than the others in as few uses of the balance as possible.

   (1) Argue that weighing 6 balls against 6 in the first weighing will not result in minimum number of weighings.
   (2) Argue that weighting 3 balls against 3 in the first weighing will also ot result in minimum number of weighings.
   (3) Come up with an optimal strategy.

2. Establish the following properties of information quantities:

   (a) [5 pts] Conditioning does not increase entropy. However, the same is not true for mutual information. Give an example of random variables X, Y, Z such that $I(X, Y|Z) > I(X, Y)$.

   (b) [5 pts] To prove Fano's inequality, we used the chain rule for entropy. Prove the following general *chain rule for entropy*:

   $$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1) \leq \sum_i H(X_i)$$

   When does equality hold in the last expression?

   (c) [5 pts] To prove the data processing inequality, we used the chain rule for mutual information. Prove the following general *chain rule for mutual information*:

   $$I(X_1, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, \ldots, X_1)$$

   Hint: Use the previous result in 2b.

(d) [5 pts] $D(p||q)$ is convex in the pair (p, q), i.e.

$$D(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1||q_1) + (1-\lambda)D(p_2||q_2)$$

Using this, argue that $H(p)$ is concave in p.

3. *Data processing inequality and sufficient statistics*

(a) [5 pts] Aruge that if $U \to X \to Y \to V$, then $I(U, V) \leq I(X, Y)$.

(Hint: You can use the data processing inequality, or alternatively use 2(c). The following fact might also be helpful: $X \to Y \to Z$ is equivalent to $Z \to Y \to X$. However, you should justify it briefly if you use it.)

(b) [5 pts] Using the definition given in class, argue that if $X_1, \ldots, X_n \overset{iid}{\sim} f_\theta = \text{uniform}(\theta, \theta+1)$, then $T(X_1, \ldots, X_n) = \{\max_i X_i, \min_i X_i\}$ is a sufficient statistic for f_θ.