10-704 Homework 4
Due: Thursday 4/23/2015

Instructions: Turn in your homework in class on Thursday 4/23/2015

1. Recall that the rate distortion function $R(D)$ can be characterised via

$$R(D) = \min_{p(t|x)} I(X;T) \quad \text{sub. to. } \mathbb{E}d(X;T) \leq D$$

   (a) Argue that R is a non-increasing in D.

   (b) Prove that R is convex in D.
   Hint: If the distributions p_1, p_2 have distortions D_1, D_2 and mutual informations I_1 and I_2, what can be said about the distortion and mutual information of a convex combination $\alpha p_1 + (1-\alpha)p_2$ of p_1 and p_2 ?

2. You have M independent channels available with (binary) capacities $C_1, \ldots, C_M$. To transmit a signal you first choose a channel and then transmit your signal over that channel. The receiver can identify which channel was used – i.e. you may assume that the output alphabets are disjoint.

   (a) Show that the capacity of this channel is $\log_2\left(\sum_i 2^{C_i}\right)$.

   (b) How would the transmitter achieve this capacity ?

3. In this problem we will derive lower bounds for parametric classification problems. In classification problems, the risk function is the expectation of the zero-one loss over the joint distribution. That is, if P_{XY} is the joint distribution, then the risk $R(f) = \mathbb{E}_{X,Y\sim P_{XY}} \mathbf{1}[f(X) \neq Y]$. We will establish minimax lower bounds on the *excess risk*, which is the risk minus the risk of the bayes optimal estimator, which is the function minimizing the risk.

   (a) Consider a classification setting where $X \in [0,1]$ and $Y|X = x \sim \text{Bernoulli}(\eta(x))$ and $\eta(x) = P(Y=1|X=x)$ is the regression function. For a distribution P over X, Y with regression function η, the Bayes optimal predictor is $f_B(x) = \mathbf{1}[\eta(x) \geq 1/2]$. Let $\mathcal{P}$ be the class of all distributions for which the Bayes optimal predictor is of the form $f_B(x) = \mathbf{1}[x \leq t]$. This means that the regression function η crosses $1/2$ at most once at some point t (and it is above half on $[0,t)$). For such a distribution, the optimal estimator is the function $\mathbf{1}[x \leq t]$ and we say that its risk is $R^\star$. Show that

$$\inf_{\widehat{f}_n} \sup_{P_{XY}\in\mathcal{P}} \mathbb{E}_{(X_i,Y_i)\sim P_{XY}}[R(\widehat{f}_n) - R^\star] \geq C\sqrt{\frac{1}{n}}$$

   You can use Le Cam's method. It is convenient to choose the two distributions to have the same marginal $P_x = \text{Uniform}[0,1]$ and only vary the conditional $P_{Y|X}$.

(b) Consider now the same set up except where the data $X \in [0,1]^d$ and the class $\mathcal{P}$ corresponds to distributions with Bayes optimal predictors of the form $f_B(x) = \prod_{j=1}^{d} \mathbf{1}[x_j \leq t_j]$. Use Fano's method to show that:

$$\inf_{\widehat{f}_n} \sup_{P_{XY} \in \mathcal{P}} \mathbb{E}_{(X_i,Y_i)\sim P_{XY}} [R(\widehat{f}_n) - R^\star] \geq C\sqrt{\frac{d}{n}}$$