Quiz 2

Date: Monday, November 21, 2016

Name: ______________________________

Andrew ID: ______________________________

Department: ______________________________

Guidelines:

1. **PLEASE DO NOT TURN THIS PAGE UNTIL INSTRUCTED.**
2. Write your name, Andrew ID, and department in the spaces provided above.
3. You have **60 minutes** for this exam.
4. This exam has **9 pages** on 9 (nine) sheets of paper, including this one.
5. This exam has a total of **50 possible points**, split between **5 "short" questions** and **2 "long" questions**. The points allocated to each question are indicated next to that question.
6. This exam is open notes. You may use any materials such as cheat sheets, class notes, etc. No electronic devices are permitted.
7. The questions vary in difficulty. The points allocated to a question do not entirely reflect its difficulty. Do not spend too much time on one question.
8. Questions only appear on one side of each sheet of paper. You may use any blank space for your answer or scratch work, but please clearly indicate your answers.

1 Short Questions

1. We wish to encode a dictionary of 4 symbols $\{a, b, c, d\}$ using a ternary alphabet $\{0, 1, 2\}$.

   (a) **[4 points]** Identify the following 4 codes as Singular (S), Non-Singular but not uniquely decodable (NS), Uniquely Decodable but not instantaneous (UD), or Instantaneous (I).

      i. $\{0, 1, 11, 21\}$

      ii. $\{01, 10, 11, 02\}$

      iii. $\{0, 1, 2, 1\}$

      iv. $\{0, 112, 11, 22\}$

   (b) **[4 points]** According to the IID source distribution

   $$p(a) = 1/3 \quad p(b) = 1/9 \quad p(c) = 2/9 \quad p(d) = 1/3$$

   (encoding based on that order) give a ternary arithmetic code for the sequence *bcd*. You may assume the decoder knows when to stop. (*Note: multiple valid answers exist; give any correctly decodable answer.*)

2. Suppose we want to transmit an input $X = (X_1, X_2)$ across two parallel Gaussian channels with joint correlation matrix Σ, under a total power constraint $\mathbb{E}\left[\|X\|_2^2\right] \leq 3$. What distribution of the input X maximizes the rate if

   (a) **[2 points]** $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

   (b) **[3 points]** $\Sigma = \begin{bmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix}$. *(Hint: Σ has unit eigenvectors $v_1 = \left(\frac{1}{\sqrt{3}}, \sqrt{\frac{2}{3}}\right)$, $v_2 = \left(\sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{3}}\right)$ and corresponding eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 0$.)*

3. **[7 points]** No justification is necessary for this problem.

   (a) For each of the following tasks, would you use the Rate-Distortion (RD) method or the Information Bottleneck (IB) method?

      i. Compress a dataset X of predictors while still being able to predict a response Y.

      ii. Compress a video without significantly sacrificing video quality.

   (b) True or False: The objective functions defining the rate-distortion function, the information bottleneck method, and the capacity of a channel can all be optimized using the Blahut-Arimoto algorithm.

   (c) Let X be a random variable, and suppose $T(X)$ is a sufficient statistic for a parameter θ. Which of the following statements is always true?

      i. $H(\theta|X) = H(\theta|T(X))$.

      ii. $H(X|\theta) \geq H(T(X)|\theta)$.

      iii. If T is minimal, then T is unique.

      iv. $P(X|T(X))$ is independent of θ.

4. [**5 points**] Consider a discrete privacy mechanism $\mathcal{M}$ over a space of data sets $\mathcal{X}$ (e.g., $\mathcal{M} : \mathcal{X} \to \mathbb{N}$). Show that, if $\mathcal{M}$ is ε-differentially private, then

$$\sup_{x,x' \in \mathcal{X} : x \sim x'} D_{KL}(\mathcal{M}(x) || \mathcal{M}(x')) \leq \varepsilon,$$

where $x \sim x'$ denotes that x and x' differ in a single entry and the KL divergence D_{KL} is over the randomness of $\mathcal{M}$.

5. **[5 points]** Consider universal prediction in the context of online linear regression. At each time point t, we

   1. observe a predictor $x_t \in \mathbb{R}^D$.
   2. output a prediction $\widehat{y}_t \in \mathbb{R}$.
   3. observe a true $y_t \in \mathbb{R}$.
   4. suffer squared error loss $\ell(y_t, \widehat{y}_t) = (y_t - \widehat{y}_t)^2$.

   Suppose we consider all linear hypotheses, i.e., our hypothesis space can be represented by $\Theta = \mathbb{R}^D$, where $w \in \Theta$ predicts $\langle w, x \rangle$ for each input $x \in \mathbb{R}^D$. Explain how to predict each $\widehat{y}_t$, using the exponential weights algorithm with a prior p_0 over Θ and learning rate η.

2 Long Questions

1. **[10 points]** Consider a channel C that takes a binary input X and returns a binary output Y, according to the following conditional distribution:

x	$\mathbb{P}[Y = 0 \mid X = x]$	$\mathbb{P}[Y = 1 \mid X = x]$
0	1	0
1	0.5	0.5

What is the capacity (in bits) of C?

Hint: Check that $I(X;Y)$ is concave in p and find the maximizer.

2. **[10 points]** Consider n IID samples $(X_1, Y_1), \ldots, (X_n, Y_n) \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

for some $\rho \in [-1, 1]$.

(a) For a fixed $\rho_* \in (0, 1)$, use Le Cam's method to derive a minimax lower bound for testing the null hypothesis $\rho = 0$ against the alternative hypothesis $|\rho| > \rho_*$.

(b) For a given constant $c \in (0, 1)$, what is largest value of ρ_* such that, according to your lower bound, every test has worst-case error probability at least c?

Hint: The KL divergence between Gaussians $\mathcal{N}(0, \Sigma_0)$ and $\mathcal{N}(0, \Sigma_1)$ over $\mathbb{R}^2$ is

$$D_{KL}(\mathcal{N}(0, \Sigma_0), \mathcal{N}(0, \Sigma_1)) = \frac{1}{2} \left(\log \frac{|\Sigma_0|}{|\Sigma_1|} - 2 + tr(\Sigma_0^{-1} \Sigma_1) \right),$$

where $|\Sigma| = \Sigma_{1,1}\Sigma_{2,2} - \Sigma_{1,2}\Sigma_{2,1}$ and $tr(\Sigma) = \Sigma_{1,1} + \Sigma_{2,2}$ denote the determinant and trace of Σ.

This space is intentionally blank. You may use it for scratch work.

Please do not mark below this line.

Problem	Max	Points
S1	8	
S2	5	
S3	7	
S4	5	
S5	5	
L1	10	
L2	10	
Total	50	