ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Information Theory and Signal Processing	Assignment date: January 17th, 2019, 16:15
Fall 2018	Due date: January 17th, 2019, 19:15

Final Exam – CE1106

There are six problems. We do not presume that you will finish all of them. Choose the ones you find easiest and collect as many points as possible. Good luck!

Name: ______________________________

Problem 1	/ 10
Problem 2	/ 20
Problem 3	/ 15
Problem 4	/ 15
Problem 5	/ 20
Problem 6	/ 20
Total	/100

Problem 1. *(Canonical Correlations)*

[10pts] Let $\mathbf{X}$ and $\mathbf{Y}$ be zero-mean real-valued random vectors with covariance matrices $R_{\mathbf{X}}$ and $R_{\mathbf{Y}}$, respectively. Moreover, let $R_{\mathbf{XY}} = \mathbb{E}[\mathbf{X}\mathbf{Y}^T]$. Our goal is to find vectors $\mathbf{u}$ and $\mathbf{v}$ such as to maximize the correlation between $\mathbf{u}^T\mathbf{X}$ and $\mathbf{v}^T\mathbf{Y}$, that is,

$$\max_{\mathbf{u},\mathbf{v}} \frac{\mathbb{E}[\mathbf{u}^T\mathbf{X}\mathbf{Y}^T\mathbf{v}]}{\sqrt{\mathbb{E}[|\mathbf{u}^T\mathbf{X}|^2]}\sqrt{\mathbb{E}[|\mathbf{v}^T\mathbf{Y}|^2]}}. \tag{1}$$

Show how we can find the optimizing choices of the vectors $\mathbf{u}$ and $\mathbf{v}$ from the problem parameters $R_{\mathbf{X}}, R_{\mathbf{Y}}$, and $R_{\mathbf{XY}}$.

Hint: Recall that we have seen in class that

$$\max_{\mathbf{v}} \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|} = \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\| = \sigma_1(A), \tag{2}$$

where $\sigma_1(A)$ denotes the maximum singular value of the matrix A. The corresponding maximizer is the right singular vector $\mathbf{v}_1$ (i.e., eigenvector of A^TA) corresponding to $\sigma_1(A)$.

Problem 2. *(Growth of Expected Capital vs Expected Growth of Capital)*

Suppose $U_1, U_2, \ldots$ are i.i.d. random variables taking values on a finite alphabet $\mathcal{U}$; let $P(u) = \Pr(U_1 = u)$ denote their common distribution. As in class let $\hat{P}_n$ denote the empirical distribution of U^n.

Suppose $f : \mathcal{U} \to [0, \infty)$ is a non-negative real valued function defined on $\mathcal{U}$. Define now the random variables $X_0, X_1, \ldots$ as $X_0 = 1$, $X_n = f(U_n)X_{n-1}$, $\forall n \geq 1$. In other words

$$X_n = \prod_{i=1}^{n} f(U_i).$$

One refers to the value $R_n = \frac{1}{n} \log X_n$ as the (exponential) *rate of growth* of X_n. (The terminology is motivated by the relationship $X_n = \exp(nR_n)$).

Fix $\alpha = \sum_u P(u) \log f(u) = E[\log f(U)]$, and for a given $\epsilon > 0$, let

$$A = \left\{Q \in \Pi : \left|\textstyle\sum_u Q(u) \log f(u) - \alpha\right| < \epsilon\right\}.$$

Let $D^* = \min_{Q \notin A} D(Q\|P)$. Observe that $D^* > 0$.

(a) [5pts] What can you say about $\Pr(|R_n - \alpha| \geq \epsilon)$ as n gets large? *Hint:* How are the events $\{|R_n - \alpha| \geq \epsilon\}$ and $\{\hat{P}_n \notin A\}$ related?

(b) [5pts] Let $\beta = \log E[f(U)]$. What is the relationship between $e_n = \frac{1}{n} \log E[X_n]$ and β? Which one of α and β is larger?

In a casino a game of chance is played. The outcome of the game is a random variable U, and if the outcome is u, the money bet on that outcome is multipled by a factor $\phi(u)$. The money bet on other outcomes is lost. The game can be played successively with independent, identically distributed outcomes.

We allocate our capital among the outcomes by placing a fraction $q(u)$ of it on outcome u. Clearly $q(u) \geq 0$ and $Q = \sum_u q(u) \leq 1$. (The fraction $1 - Q$ is the fraction of our capital not bet on the game and kept in reserve.) Observe that $f(u) = (1 - Q) + q(u)\phi(u)$ is the factor our capital is multipled by if the outcome of the game is u.

Let $X_0 = 1$ be our initial capital, and let X_n, $n = 1, 2, \ldots$ denote our capital as we play the game repeatedly with a fixed allocation strategy q.

(c) [5pts] Suppose $\mathcal{U} = \{0, 1\}$, $P(0) = 1/4$, $P(1) = 3/4$, $\phi(0) = \phi(1) = 2$. What is the allocation q that maximizes the value of β in (b)?

(d) [5pts] Continuing with (c) and the allocation you just found, what is the value of α? What will happen to our capital X_n in the long run if we repeatedly play the game?

Problem 3. *(Hypothesis Testing and Exponential Families)*

Let P denote the zero-mean and unit-variance Gaussian distribution. Assume that you are given N iid samples distributed according to P and let $\hat{P}_N$ be the empirical distribution.

Let Π denote the set of distributions with second moment $\mathbb{E}[X^2] = 2$. We are interested in

$$\lim_{N\to\infty} \frac{1}{N} \log \Pr\{\hat{P_N} \in \Pi\} = -\inf_{Q\in\Pi} D(Q\|P).$$

1. [10pts] Determine $-\text{arginf}_{Q\in\Pi} D(Q\|P)$, i.e., determine the element Q for which the infinum is taken on.
2. [5pts] Determine $-\inf_{Q\in\Pi} D(Q\|P)$.

Problem 4. *(Choose the Shortest Description)*

Suppose $\mathcal{C}_0 : \mathcal{U} \to \{0,1\}^*$ and $\mathcal{C}_1 : \mathcal{U} \to \{0,1\}^*$ are two prefix-free codes for the alphabet $\mathcal{U}$. Consider the code $\mathcal{C} : \mathcal{U} \to \{0,1\}^*$ defined by

$$\mathcal{C}(u) = \begin{cases} 0\mathcal{C}_0(u) & \text{if length}\mathcal{C}_0(u) \le \text{length}\mathcal{C}_1(u) \\ 1\mathcal{C}_1(u) & \text{else.} \end{cases}$$

Observe that $\text{length}(\mathcal{C}(u)) = 1 + \min\{\text{length}(\mathcal{C}_0(u)), \text{length}(\mathcal{C}_1(u))\}$.

(a) [5pts] Is $\mathcal{C}$ a prefix-free code? Explain.

(b) [5pts] Suppose $\mathcal{C}_0, \dots, \mathcal{C}_{K-1}$ are K prefix-free codes for the alphabet $\mathcal{U}$. Show that there is a prefix-free code $\mathcal{C}$ with

$$\text{length}(\mathcal{C}(u)) = \lceil \log_2 K \rceil + \min_{0 \le k < K-1} \text{length}(\mathcal{C}_k(u)).$$

(c) [5pts] Suppose we are told that U is a random variable taking values in $\mathcal{U}$, and we are also told that the distribution p of U is one of K distributions $p_0, \dots, p_{K-1}$, but we do not know which. Using (b) describe how to construct a prefix-free code $\mathcal{C}$ such that

$$E[\text{length}(\mathcal{C}(U))] \le \lceil \log_2 K \rceil + 1 + H(U).$$

[Hint: From class we know that for each k there is a prefix-free code $\mathcal{C}_k$ that descibes each letter u with at most $\lceil -\log_2 p_k(u) \rceil$ bits.]

Problem 5. *(Inner Products)*

Consider the standard n-dimensional vector space $\mathbb{R}^n$.

1. [5pts] Characterize the set of matrices W for which $\mathbf{y}^T W \mathbf{x}$ is a valid inner product for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

2. [5pts] Prove that *every* inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ on $\mathbb{R}^n$ can be expressed as $\mathbf{y}^T W \mathbf{x}$ for an approriately chosen matrix W.

3. [10pts] For a subspace of dimension $k < n$, spanned by the basis $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k \in \mathbb{R}^n$, express the orthogonal projection operator (matrix) with respect to the general inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T W \mathbf{x}$. *Hint:* For any vector $\mathbf{x} \in \mathbb{R}^n$, express its projection as $\widehat{\mathbf{x}} = \sum_{j=1}^{k} \alpha_j \mathbf{b}_j$.

Problem 6. *(Thompson Sampling with Bernoulli Losses)*

This problem deals with a Bayesian approach to multi-arm bandits. Although we will not pursue this facet in the current problem, the Bayesian approach is useful since within this framework it is relatively easy to incorporate prior information into the algorithm.

Assume that we have K bandits, and that bandit k outputs a $\{0,1\}$-valued Bernoulli random variable with parameter $\theta_k \in [0,1]$. Let π be the uniform prior on $[0,1]^K$, i.e., the uniform prior on the set of all parameters $\theta = (\theta_1, \cdots, \theta_K)$. Let

$$T_k^1(t) = |\{\tau \leq t : A_\tau = k; Y_\tau = 1\}|,$$
$$T_k^0(t) = |\{\tau \leq t : A_\tau = k; Y_\tau = 0\}|.$$

In words, $T_k^1(t)$ is the number of times up to and including time t that we have chosen action k and the output of arm k was 1 and similarly $T_k^0(t)$ is the number of times up to and including time t that we have choses action k and the output of the arm k was 0.

The goal is to find the arm with the highest parameter, i.e., the goal is to determine

$$k^* = \text{argmax}_k \theta_k.$$

In the Bayesian approach we proceed as follows. At time time t:

1. Compute for each arm k the distribution $p(\theta_k(t)|T_k^1(t-1), T_k^0(t-1))$.
2. Generate samples of these parameters according to their distributions.
3. Pick the arm j with the largest sample.
4. Observe the output of the j-th arm, call it $Y_j(t)$, and update the counters T_j^1 and T_j^0 accordingly.

Show that this algorithm "works" in the sense that eventually it will pick the best arm. More precisely, show the following two claims.

1. [10pts] Show that $p(\theta_k(t)|T_k^1(t-1), T_k^0(t-1))$ is a Beta distributed and determine α and β.
2. [10pts] Show that as t tends to infinity the probability that we choose the correct arm tends to 1. [HINT: To simplify your life, you can assume that for every arm k, $T_k^1(t-1) + T_k^0(t-1) \overset{t\to\infty}{\to} \infty$.]

NOTE: Recall that the density of the Beta distribution on $[0,1]$ with parameters α and β is equal to

$$f(x; \alpha, \beta) = \text{constant } x^{\alpha-1}(1-x)^{\beta-1}.$$

Further, the expected value of $f(x; \alpha, \beta)$ is $\frac{\alpha}{\alpha+\beta}$ and its variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.