Problem Set 6

Exponential Families & Linear Algebra

For the Exercise Session of Nov 17 - Nov 24 — **Due: Tue, December 1, 10am, on Moodle**

1 The Homework

The following two problems are the core of this homework. You work out solutions and turn them in. Problem 1 is suitable for the exercise session November 17. Problem 2 is suitable for November 24.

Problem 1: The Maximum Likelihood Estimator and I-projections

Let X^n be i.i.d. random variables taking values in a finite set $\mathcal{X}$ of size k. Recall that when restricted to distributions from a set $\mathcal{P}$, the maximum likelihood estimator $: \mathcal{X}^n \to \mathcal{P}$ is given as

$$P^*_{\text{MLE}} = (x^n) = \arg\max_{P \in \mathcal{P}} P\left(X^n = x^n\right).$$

Note that the two parts of this problem are independent of each other.

(i) Show that the maximum likelihood estimate P^*_{MLE} is equal to the I-projection

$$P^* := \arg\min_{P \in \mathcal{P}} D\left(\hat{P} \middle\| P\right)$$

where $\hat{P}$ denotes the empirical distribution of x^n.

(ii) Let P_0 be a distribution over $\mathcal{X}$ and let $\mathcal{P}$ be defined as the minimal exponential family

$$\mathcal{P} := \left\{P : P(x) = P_0(x)e^{-\langle\theta,\phi(x)\rangle - A(\theta)}, \theta \in \Theta\right\}$$

for some open parameter space $\Theta \subseteq \mathbb{R}^d$, sufficient statistic ϕ, and normalization function A. Then show that the maximum likelihood estimate P^*_{MLE} is the I-projection P^* of P_0 onto the linear family

$$\mathcal{L}(x^n) := \left\{P : E_P\left[\phi(X)\right] = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\right\}.$$

In other words, show that

$$\arg\min_{P \in \mathcal{L}(x^n)} D\left(P_0 \| P\right) = \arg\max_{Q \in \mathcal{P}} Q(X^n = x^n).$$

Problem 2: Canonical Correlation Analysis

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 for the singular value decomposition 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 $A^T A$) corresponding to $\sigma_1(A)$.

2 Additional Problems

For the next five problems, you do not need to turn in solutions. For the exam, we do expect you know how to solve these problems. So, you may keep these problems around and tackle them in preparation for the Midterm and/or Final exam. Or you may solve them now.

Problem 3: Maximum Entropy of Sum of Random Variables

Let $Y = X_1 + X_2$. Find the maximum entropy of Y under the constraint $\mathbb{E}[X_1^2] = P_1$, $\mathbb{E}[X_2^2] = P_2$:

(a) If X_1 and X_2 are independent.

(b) If X_1 and X_2 are allowed to be dependent.

Problem 4: Exponential Families and Conjugate Priors

Let $p_\theta(x) = h(x)e^{\langle \phi(x), \theta \rangle - A(\theta)}$ denote a generic exponential family with sufficient statistics $\phi(x)$ and parameter θ.

Assume that we receive iid samples from this family, call them $\{x_i\}_{i=1}^n$. From these samples, we want to infer the unknown parameter θ via a maximum a-posteriori (MAP) procedure.

In order to apply a MAP procedure we need to define a prior distribution on the parameter θ. Consider the family of prior distributions $q_{\mu,\lambda}(\theta) = K(\mu,\lambda)e^{\langle \theta, \mu \rangle - \lambda A(\theta)}$, parametrized by (μ, λ). Note that this is also an exponential family. However, we have written it in a slightly non-standard form, where $K(\mu,\lambda)$ denotes the normalization constant which is a function of the parameters (μ, λ).

(i) Write down the posterior distribution $p_{\mu,\lambda}(\theta \mid x_1, \cdots, x_n)$ for a fixed set of parameters (μ, λ).

(ii) If you have not already done so in part (i), write the posterior as explicitly and compactly as you can. Justify why we called $q_{\mu,\lambda}(\theta)$ a conjugate prior.

(iii) Derive the MAP estimator of the parameter θ given the samples $\{x_i\}_{i=1}^n$ starting with the posterior derived in (ii). When will the estimate be unique?

Problem 5: Maximum Entropy for Cartesian Product of Sets

What is the maximum entropy distribution, call it $p(x, i)$, on $[0, \infty] \times \mathbb{N}$, both of whose marginals have mean $\mu > 0$. (I.e., in one axis the distribution is over the positive reals, whereas in the other one it is over the natural numbers.)

Problem 6: Exponential Families and Maximum Entropy

For $t > 0$, consider a family of distributions supported on $[t, +\infty]$ such that $\mathbb{E}[\ln X] = \frac{1}{\alpha} + \ln t$, $\alpha > 0$.

1. What is the parametric form of a maximum entropy distribution satisfying the constraint on the support and the mean?

2. Find the exact form of the distribution.

Problem 7: Dual Representation of Norm

(a) Assume that $p > 0$ and $q > 0$ satisfy $1/p + 1/q = 1$. Show that the following inequality holds for all $a \geq 0$ and $b \geq 0$.

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \tag{3}$$

Show that the equality holds if $a^p = b^q$.

Hint: Use the concavity of log function

(b) Given vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ show that,

$$\frac{\sum_{i=1}^n |x_i y_i|}{\|\mathbf{x}\|_p \|\mathbf{y}\|_q} \leq 1 \tag{4}$$

What is the condition for equality?

(c) Show that

$$||\mathbf{x}||_p = \sup_{\substack{\mathbf{y} \in \mathbb{R}^n: \\ \|\mathbf{y}\|_{p^*} = 1}} \langle \mathbf{y}, \mathbf{x} \rangle.$$

where $1/p + 1/p^* = 1$.