HOMEWORK 1 - V3

CSC2547/STA4273 WINTER 2019

University of Toronto

VERSION HISTORY:	V0 $\to$ V1:	FIX TRANSPOSE (Q1.2), FIX NORM (Q1.3)
	V1 $\to$ V2:	FIX STATEMENT (Q1.2), ADD RANGE OF ϕ (Q2.3)
	V2 $\to$ V3:	ADD PART B (Q1.2), CLARIFY GRADIENTS
	V3 $\to$ V4:	CLARIFY (Q2.1)

1. Gaussian mean estimation.

1.1. *Optimal shrinkage factor [10pts].* Let $X_1, X_2, ..., X_n \in \mathbb{R}^d$ be i.i.d. multivariate Gaussian random vectors, i.e., $X_i \sim \mathcal{N}(\mu, \sigma^2 I)$. Denoting the sample mean estimator with $\hat{\mu} \triangleq \frac{1}{n}\sum_{i=1}^n X_i$, consider an estimator of the form $\hat{\mu}^s = \left(1 - \frac{\tau}{\|\hat{\mu}\|_2^2}\right)\hat{\mu}$. Find the optimal τ that minimizes the risk $R(\hat{\mu}^s, \mu) = \mathbb{E}[\|\hat{\mu}^s - \mu\|_2^2]$.

1.2. *Generalizing Stein's lemma [10pts].* Let $X \sim p_\eta(x)$ and $g_\eta : \mathbb{R}^d \to \mathbb{R}^d$ where $p_\eta(x)$ and $g_\eta(x)$ are differentiable w.r.t η and x, and let $\mathbb{E}[g_\eta(X)] = \xi(\eta)$ for some function ξ. Show that

(a) $\mathbb{E}[\nabla_x \log p_\eta(X) g_\eta(X)^\top] + \mathbb{E}[\nabla_x g_\eta(X)] = 0$,
(b) $\mathbb{E}[\nabla_\eta \log p_\eta(X) g_\eta(X)^\top] + \mathbb{E}[\nabla_\eta g_\eta(X)] = \nabla_\eta \xi(\eta)$.

1.3. *Generalizing SURE [10pts].* Let $X \sim \mathcal{N}(\mu, \Sigma)$ where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$. If $\hat{\mu}(X) \in \mathbb{R}^d$ is an estimator of the form $X + g(X)$ where $g : \mathbb{R}^d \to \mathbb{R}^d$ is differentiable. Define the functional

$$S(X, \hat{\mu}) = \mathrm{Tr}(\Sigma) + 2\,\mathrm{Tr}(\Sigma \nabla_x g(X)) + \|g(X)\|_2^2.$$

Then show that $S(X, \hat{\mu})$ is an unbiased estimator of the risk, i.e., $\mathbb{E}[\|\hat{\mu}(x) - \mu\|_2^2] = \mathbb{E}[S(X, \hat{\mu})]$.

2. Exponential families.

2.1. *Second moment [10pts].* For a random variable $X \sim p_\eta(x) = \exp(\langle \eta, \phi(x)\rangle - \psi(\eta))$, let $\mathbb{E}[\phi(X)] = \mu$. For $\xi \in \mathbb{R}^d$, find $\mathrm{Tr}(\mathbb{E}[(\phi(X) - \xi)(\phi(X) - \xi)^\top])$ in terms of ξ and $\nabla_\eta^i \psi(\eta)$ for $i \geq 0$.

2.2. *Score function [10pts].* Assume that $X \sim p_\eta(x)$ where p_η is not necessarily in the exponential family form. Denote the log-likelihood by $\ell_\eta(x) = \log p_\eta(x)$, show that

(a) $\mathbb{E}[\nabla_\eta \ell_\eta(X)] = 0$.
(b) $\mathbb{E}[\nabla_\eta \ell_\eta(X) \nabla_\eta \ell_\eta(X)^\top] = -\mathbb{E}[\nabla_\eta^2 \ell_\eta(X)]$ (Problem 1.2 may be helpful).

2.3. *Maximum entropy principle [Bonus 2pts].* Assume that $p(x)$ is a probability mass function of a discrete random variable taking values from a finite set $\mathcal{X}$. Entropy of p is defined as $H(p) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$. For $\phi : \mathcal{X} \to \mathbb{R}^d$, show that the maximum entropy distribution satisfying $\mathbb{E}_p[\phi(X)] = \mu \in \mathbb{R}^d$ is a member of exponential family. That is, show that the solution to

$$\underset{p}{\text{maximize}}\, H(p) \quad \text{subject to: } \mathbb{E}_p[\phi(X)] = \mu,$$

is an exponential family. (Hint: Write the Lagrangian associated with the above optimization problem. Since $\mathcal{X}$ is finite, think of $p(x)$ as a vector and maximize over it.)