HOMEWORK 0

CSC2547/STA4273 WINTER 2019

University of Toronto

1. Calculus.

1.1. *Directional derivative and gradient.* For a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ and a unit vector $u \in \mathbb{R}^d$ with $\|u\|_2 = 1$, directional derivative is given as $\nabla_u f(x) = \lim_{h\downarrow 0}(f(x + hu) - f(x))/h$. Show that $\nabla_u f(x) = \langle \nabla f(x), u\rangle$. What does the gradient of f represent?

1.2. *Generalized linear models.* For a random data sample pair (y, x), population version of a (canonical) GLM loss function looks like

$$\ell(\beta) = \mathbb{E}[\Psi(\langle x, \beta\rangle) - y\langle x, \beta\rangle].$$

Show that $\ell(\beta)$ is always convex if Ψ is convex.

2. Linear Algebra.

2.1. *Dual norm of ℓ_2.* For a given norm $\|\cdot\|$ and a vector $x \in \mathbb{R}^d$, dual norm is defined as $\|x\|_* = \sup_{\|u\|\le 1} \langle u, x\rangle$. Find the dual norm of ℓ_2-norm (Here $\langle u, x\rangle = u^\top x$.)

2.2. *Dual norm of ℓ_1.* Using the definition in 2.1, find the dual norm of ℓ_1-norm.

2.3. *Expectation of ℓ_2-norm.* For a random vector $x \in \mathbb{R}^d$ with $\mathbb{E}[x] = \mu$ and $\mathrm{Var}(x) = \Sigma$, give an expression for $\mathbb{E}[\|x\|_2^2]$ in terms of μ and Σ (Start with the definition of Var and use trace.)

3. Probability.

3.1. *Law of iterated expectation.* For random variables X, Y and Z, show that

$$\mathbb{E}[\mathbb{E}[X|Y, Z]|Z] = \mathbb{E}[X|Z].$$

3.2. *Integration by parts.* For a non-negative random variable X, show that

$$\int_0^\infty \mathbb{P}(X > t)dt = \mathbb{E}[X].$$

What condition do you need?

3.3. *Total variation distance.* Total variation distance between two probability measures p, q on a countable set Ω is defined as

$$d_{\mathrm{TV}}(p, q) = \sup_{A\subset\Omega} |p(A) - q(A)|.$$

Show that $d_{\mathrm{TV}}(p, q) = \frac{1}{2}\|p - q\|_1$ where $\|p - q\|_1$ is defined as

$$\|p - q\|_1 = \sum_{w\in\Omega} |p(w) - q(w)|.$$