Homework 7
CS-526 Learning Theory

1. Short problems

1. [Several correct answers possible.] Let $(x_i, y_i) \in \mathbb{R} \times \{0,1\}$ for $i \in \{1, \ldots, n\}$. Let $\hat{y}_i(w) = 1/(1+e^{-wx_i})$. Define

$$f : w \in \mathbb{R} \mapsto -\sum_{i=1}^{n} \left[y_i \log\left(\hat{y}_i(w)\right) + (1-y_i)\log\left(1-\hat{y}_i(w)\right)\right] + \lambda|w|,$$

where $\lambda > 0$. The function f is:

   (a) convex.
   (b) differentiable everywhere.
   (c) subdifferentiable everywhere.
   (d) Lipschitzian.

2. Consider the function

$$f(x) = x^2 + 2.5\cos x + |x|,$$

defined on the real line $\mathbb{R}$. Which of the following statements is correct and why/why not? The function f is:

   (a) convex
   (b) differentiable everywhere
   (c) subdifferentiable everywhere

3. Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a differentiable Lipschitz function with constant ρ. Define $h_\alpha : \mathbb{R}^d \mapsto \mathbb{R}$, with $h_\alpha(x) = g(\|x\|^\alpha)$ where $\alpha > 0$. For which values of $\alpha > 0$ can we conclude that h_α a Lipschitz function without further information on g ? Give a Lipschitz constant when this is the case.

4. Let $f(x) = a|x|^3 + b|x| + c$ for $a, b \in \mathbb{R}_+$ and $c \in \mathbb{R}$. Is this function convex ? If yes what are the subgradient sets $\partial f(x)$?

5. Let $G(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$ and the convolution $f_G(x) = \int_{\mathbb{R}} dz G(z-x) f(z)$. Consider the standard Gaussian random variable $Z \sim \mathcal{N}(0,1)$. Consider the random map $x \mapsto Zf(x+Z)$. Which is true ?

   (a) This random map is a stochastic gradient of f_G.

(b) This random map cannot be a stochastic gradient since it does not contain any derivative.

2. Gradient Descent for Positive Semi-definite Matrices

Let $X, Y \in \mathbb{R}^{n\times n}$ be $n \times n$ real matrices and $A, B \in \mathbb{R}^{n\times n}$ be $n \times n$ real symmetric and positive definite matrices. Let $F : \mathbb{R}^{n\times n} \mapsto \mathbb{R}$ the function $F(X) = \frac{1}{2}\mathrm{Tr}X^T BX$.

1. Show that $F(X) \geq 0$ for any X.
2. Compute the second derivative of
$$f(s) = \mathrm{Tr}(sX^T + (1-s)Y^T)B(sX + (1-s)Y)$$
for $s \in [0, 1]$ and deduce that F is a convex function.
3. Deduce the inequality $F(Y) - F(X) \geq \mathrm{Tr}X^T B(Y - X)$. Is F Lipschitz ?
4. Consider now the function $G : \mathbb{R}^{n\times n} \mapsto \mathbb{R}$ with $G(X) = \frac{1}{2}\mathrm{Tr}(X - I)^T A(X - I)$ where I is the identity matrix. Define $L(X) = F(X) + G(X)$.
   (a) Write down the gradient descent algorithm for L. Call X_t the updated matrix at time t.
   (b) Assume that the operator norm $\|X_t\| \leq M$ stays bounded uniformly in n. Show that
$$\|\frac{1}{T}\sum_{t=1}^{T} X_t - (B + A)^{-1}A\| \leq \frac{2M}{\eta T}\|(B + A)^{-1}\|$$