Assignment cum Practice Questions (Set 1):

1. Let $f : C \to \Re$ be a differentiable convex function on the convex set *C*. If $\nabla f(x)$ is *L*-Lipscitz, prove that: $\frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \le f(y) - f(x) - (x-y)^T \nabla f(x) \le \frac{L}{2}\|y-x\|^2$.

2. Let $f : \Re^n \to \Re$ be a convex differentiable function with *L*- Lipschitz continuous first derivatives, with $L > 0$. Suppose that f has a minimizer on R^n and consider the following gradient-descent updates from a given x_0:

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k),\ k \ge 0.$$

3. Prove that the minimizing sequence $\{x_k\}_0^\infty$ converges to a minimizer x^* of f (as defined in Q. 2) and we have the following global convergence rate estimate:

$$f(x_k) - f(x^*) \le \frac{5L}{2\sum_{i=0}^{k-1}\|x_i - x^*\|^{-2}}.$$

6. Let $f : C \to \Re$ be a differentiable and strictly convex function on the convex set *C*. If $\nabla f(x)$ is *L*-Lipscitz, prove that the following function (a Lyapunov Energy function for Gradient Descent):

$$L(x) = \frac{1}{2}\|x - x^*\|^2$$

satisfies the inequality: $L(x_{k+1}) \le L(x_k)$ where x_k is the *k*-th iterate on a Gradient Descent (GD) iteration on the function *f*(*x*). For this proof, do we need any assumption on the learning rate of the GD? Clarify.

7. (Divergence between multivariate normal distributions): Let P_1 be $N(\theta_1, \Sigma)$ and P_2 be $N(\theta_2, \Sigma)$, where $\Sigma \succ 0$ is a positive definite matrix. What is $D_{KL}(P_1||P_2)$?

8. (Mixtures are as good as point distributions): Let *P* be a Laplace (λ) distribution on ℝ, meaning that $X \sim P$ has density

$$p(x) = \frac{\lambda}{2}\exp(-\lambda|x|).$$

Assume that $X_1, \ldots, X_n \overset{i.i.d}{\sim} P$, and let P^n denote the *n*-fold product of *P*. In this problem, we compare the predictive performance of distributions from the normal location family $\mathcal{P} = \{N(\theta, \sigma^2) : \theta \in R\}$ with the mixture distribution Q^π over P defined by the normal prior distribution $N(\mu, \tau^2)$, that is, $\pi(\theta) = (2\pi\tau^2)^{-1/2}\exp(-(\theta - \mu)^2/2\tau^2)$.

a) Let $P_{\theta,\Sigma}$ be the multivariate normal distribution with mean $\theta \in R^n$ and covariance $\Sigma \in R^{n\times n}$. What is $D_{KL}(P^n \| P_{\theta,\Sigma})$?
b) Show that $inf_{\theta\in R^n} D_{KL}(P^n \| P_{\theta,\Sigma}) = D_{KL}(P^n \| P_{0,\Sigma})$, that is, the mean-zero normal distribution has the smallest KL-divergence from the Laplace distribution.

9. (Generalized "log-sum" inequalities): Let $f : \mathbb{R}_+ \to \mathbb{R}$ be an arbitrary convex function.

a) Let a_i, b_i, $i = 1, \ldots, n$ be non-negative reals. Prove that

$$\left(\sum_{i=1}^{n} a_i\right) f\left(\frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} a_i}\right) \leq \sum_{i=1}^{n} a_i f\left(\frac{b_i}{a_i}\right).$$

b) Generalizing the preceding result, let $a : \mathcal{X} \to \mathbb{R}_+$ and b : $\mathcal{X} \to \mathbb{R}_+$, and let $u : \mathcal{X} \to \mathbb{R}_+$ satisfy $\int u(x)dx < \infty$. Show that

$$\int a(x)u(x)dx\, f\left(\frac{\int b(x)u(x)dx}{\int a(x)u(x)dx}\right) \leq \int a(x) f\left(\frac{b(x)}{a(x)}\right) u(x)dx.$$

(Hint: use (after proving) the fact that the perspective of a function f, defined by $h(x, t) = tf(x/t)$ for $t > 0$, is jointly convex in x and t)