Problems 1

1. Here's the cost function of the linear regression. (1 pts)

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^\top x^{(i)} - y^{(i)})^2$$

   (a) Show that following function $J(\theta)$ can be written in a quadratic form, which is convex. (0.5 pts)

   (b) Here's the definition of a convex function. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain is a convex set and for all x, y in its domain, and all $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

   Show that $J(\theta)$ is a convex function by using this definition. (0.5 pts)

2. Here's an update rule for Newton's method. (1 pts)

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - H^{-1}\nabla_\theta J(\theta^{(t)}),$$

, where H is the Hessian matrix of a least square problem.

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^\top x^{(i)} - y^{(i)})^2.$$

Show that one iteration of Newton's method gives us the solution to the least squares problem.

3. Positive definite matrices. (2 pts)

A matrix $A \in \mathbb{R}^{n\times n}$ is *positive semi-definite* (PSD), denoted $A \succeq$, if $A = A^\top$ and $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$. matrix A is *positive definite*, denoted $A \succ 0$, if $A = A^\top$ and $x^\top Ax > 0$ for all $x \neq 0$.

   (a) Show that the identity matrix I is positive definite matrix. (0.5 pts)

   (b) Let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^\top$ is positive semidefinite. (0.5 pts)

   (c) Let $A \in \mathbb{R}^{n\times n}$ be positive semidefinite and $B \in \mathbb{R}^{m\times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is $BAB^\top$ PSD? If so, prove it. If not, give a counter example with explicit A,B. (0.5 pts)

(d) Consider a loss function for logistic regression.

$$J(\theta) = -\sum_{i=1}^{N} y^{(i)} \log(\sigma(\theta^\top x^{(i)})) + (1 - y^{(i)}) \log(1 - \sigma(\theta^\top x^{(i)})$$

Show that the Hessian H of the cost function is positive semidefinite, which implies that the cost function above is convex. (0.5 pts)

Hint: Compute Hessian for both $y^{(i)} = 0$ and $y^{(i)} = 1$ cases separately. Also note that $\nabla_\theta \sum = \sum \nabla_\theta$.

4. Linear regression (coding problem). (6 pts)

You should submit the codes as *Jupyter notebook, .ipynb file* or *google colab link*. All codes should be executable.

(a) Consider a linear function $f(x, y) = 2x + 3y + 1$. Let's generate 100 data examples $\{(x, y, z)\}$ by uniformly sampling from domains $x \in [-1, 1]$ and $y \in [-1, 1]$. By adding gaussian noise, $\epsilon \sim N(0, I)$, we can generate $z = 2x + 3y + 1 + \epsilon$. Provide a plot of the generated data in 3 dimensional space. (0.5 pts)

(b) We build a parameterized linear model with $\theta \in \mathbb{R}^3$, $f_\theta(x, y) = \theta_0 x + \theta_1 y + \theta_2$. Solve a least square problem by using a normal equation. Provide an optimal solution θ and draw the obtained hyperplane in 3d dimensional space with the generated training data. (0.5 pts)

(c) Solve a least square problem by using a gradient descent algorithm. Provide a training curve plot (x-axis: the number of iterations, y-axis: loss values) and the optimal solution. Try 5 different learning rates and draw all training curves in a plot. Compare to the solution with the solution by normal equation. (1.0 pts)

(d) Sample 10 θ from a prior distribution $N(0, I)$. Draw the hyperplanes based on the sampled θ with the generated training data. (1.0 pts)

(e) Solve a regularized least square problem by minimizing objective function, $J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta_0 x^{(i)} + \theta_1 y^{(i)} + \theta_2 - z^{(i)})^2 + \lambda||\theta||_2^2$. Try 5 different *lambda* and draw all solution for each *lambda* with the generated training data. Using both a closed form solution, $(X^\top X - \lambda I)^{-1} X^\top y$ and gradient descent algorithm and compare the optimal solutions between these two. (1.0 pts)