Math for Deep Learning - Homework 04

Name:____

1. Convex Optimization. Here just undrstand the picture.
2. Use Lagrange Multipliers to
   (a) Minimize $f(x, y) = x^2 + y^2$ on the hyperbola $xy = 1$.
   (b) Maximize $f(x, y, z) = 2x + 3y + 5z$ on the sphere $x^2 + y^2 + z^2 = 1$.
3. Use gradient descent to compute one extremma (you can use a spreadsheet or python here). How many steps were required for each?
   (a) $f(x) = x^3 - 4x$ use $x_1 = 1$, $\epsilon = 0.01$, $\eta = 0.01$
   (b) $f(x) = x^3 - 4x$ use $x_1 = -1$, $\epsilon = 0.01$, $\eta = 0.01$
   (c) $f(x, y) = x^2 - 3x + y^2$ use $\mathbf{u_1} = \langle 1, 1 \rangle$, $\epsilon = 0.01$, $\eta = 0.01$
4. Use gradient descent with momentum to compute one extremma (you can use a spreadsheet or python here). Compare steps to Problem 3.
   (a) $f(x) = x^3 - 4x$ use $x_1 = 1$, $\epsilon = 0.01$, $\eta = 0.01$, $\alpha = 0.02$
   (b) $f(x) = x^3 - 4x$ use $x_1 = -1$, $\epsilon = 0.01$, $\eta = 0.01$, $\alpha = 0.02$
   (c) $f(x, y) = x^2 - 3x + y^2$ use $\mathbf{u_1} = \langle 1, 1 \rangle$, $\epsilon = 0.01$, $\eta = 0.01$, $\alpha = 0.02$
5. For the following data, we are trying to approximate the best line. That is, we would like a function $f(x) = ax + b$ where our error is minimal. Our goal is to find a and b. Here we will use Stochastic Gradient descent and use the following loss function

$$L(x) = (y - \hat{f}(x))^2 = (y - (ax + b))^2 = .$$

$$\frac{\partial}{\partial a} = -2x(y - ax - b)$$

$$\frac{\partial}{\partial b} = -2(y - ax - b)$$

So the stochastic gradient is

$$\left\langle \frac{\partial}{\partial a}, \frac{\partial}{\partial b} \right\rangle = \langle -2x(y - ax - b), = -2(y - ax - b) \rangle. \tag{1}$$

Write your own python program using a starting point of $\mathbf{u_1} = \langle a, b \rangle = \langle 1, 1 \rangle$, $\epsilon = 0.01$, $\eta = 0.01$ for the data

$(0, 4)$, $(1, 3)$, $(2, 3)$ and $(3, 0)$.

So your program would do the following.

- Select a point at random.
- Compute the gradient using equation 1.
- Update a, and b as in the regular gradient descent method.
- Compare. If your new a and b differ from the previous a, b by more than ϵ.
- If the diffference is more than ϵ, repeat. If the difference is less than ϵ stop. That is your answer.