MSC FUNDAMENTAL PRINCIPLES OF DATA SCIENCE

THEORY EXERCISES: UNCONSTRAINED OPTIMIZATION

Authors:

Jordi Segura

26 October 2022

1 Exercise 3

a) **Investigate whether the origin is an extremum of the function**

From the function:

$$f(x_1,x_2,x_3) = \alpha x_1^2 \exp x_2 + x_2^2 \exp x_3 + x_3^2 \exp x_1$$

We have to investigate if the origin: (0,0,0) is an extremum point. To do this, we should compute the gradient and look at these coordinates. Furthermore, the Hessian is necessary in order to ensure it is a maximum or a minimum.

$$\nabla f(x_1,x_2,x_3) = \left[2\alpha x_1 \exp x_2 + x_3^2 \exp x_1, \alpha x_1 \exp x_2 + 2x_2 \exp x_3, x_2^2 \exp x_3 + 2x_3 \exp x_1\right]$$

The Hessian $\nabla^2 f$ in this case with x_1, x_2, x_3 is defined as:

$$\nabla^2 f(x_1,x_2,x_3) = \begin{bmatrix} 2\alpha \exp x_2 + x_3^2 \exp x_1 & 2\alpha x_1 \exp x_2 & 2x_3 \exp x_1 \\ 2\alpha x_1 \exp x_2 & \alpha x_1 \exp x_2 + 2\exp x_3 & 2x_2 \exp x_3 \\ 2x_3 \exp x_1 & 2x_2 \exp x_3 & x_2^2 \exp x_3 + 2\exp x_1 \end{bmatrix}$$

From where we can first substitute for our coordinate (0,0,0) and get:

$$\nabla^2 f(0,0,0) = \begin{bmatrix} 2\alpha & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Intuitively, we see that we will find a minima if $\alpha > 0$, otherwise, the matrix is non-definite and a further analysis should be done, not in the scope of this exercise!

2 Exercise 10

a)

$$\max_{(x_1,x_2)\in\Re^2} f(x_1,x_2)$$

subject to $g(x_1,x_2) = c$**, where** $c \in \Re$ **is a parameter. Suppose the problem admits a solution** $(x^*(c),\lambda^*(c))$ **for every c (at least in some open interval for c). Define** $f^*(c) = f(x_1^*(c),x_2^*(c))$**. Prove that:**

$$\frac{df^*(c)}{dc} = \lambda^*(c)$$

First of all, we must rewrite the problem as a minimization in order to find the minima:

$$\min_{(x_1,x_2)\in\Re^2} -f(x_1,x_2),\ w.r.t\ g(x_1,x_2) = c$$

. Is important to tweak the constraint in order to have something equal to 0 to work with and the c as a free variable:

$$g'(x_1,x_2) = g(x_1,x_2) - c = 0$$

Once here, from definition in the problem of $f^*(c)$ we can also write g w.r.t (c) as:

$$g^*(c) = g(x_1^*(c), x_2^*(c)) - c$$

Then, using the definition of Lagrange multipliers, we may proceed to prove it using the similarity between $f^*(c)$ and $g^*(c)$. Before, remember that x^* is a global minima, so the gradient is equal to 0:

$$\nabla f(x^*) = \lambda \nabla g(x^*), \nabla f(x^*) = 0, \nabla g(x^*) = 0$$

$$\frac{df^*(c)}{dc} = f_{x1}(x^*)\frac{x_1}{dc} + f_{x2}(x^*)\frac{x_2}{dc}$$

$$= -\lambda(c) g_{x1}'(x^*)\frac{x_1}{dc} - \lambda(c) g_{x2}'(x^*)\frac{x_2}{dc} = -\lambda(c)(g_{x1}'(x^*)\frac{x_1}{dc} + g_{x2}'(x^*)\frac{x_2}{dc}) = -\lambda(c)\frac{dg^*(c)}{dc} = \lambda(c)$$

Where $\frac{dg^*(c)}{dc}$ must be -1 due to deriving $-c$ and the fact that x^* is a minima, so equal to 0. Hence, it is proved.

3 Exercise 13

$$f(x) = \sum_{i=1}^{m} w_i ||x^* - y_i||$$

a)

As we know, the summatory of convex functions results in a convex function. Hence, the problem remains - once we have said the norm of the distance between our central point and the weight points is convex - in demonstrate if the minimum is global.

In the visual representation, we can see how an stable system like this with a fixed weights, the only way to minimize the moment is trying to reduce the distance between the central point, which is not fixed and has to be moved to help to minimize this equation, and the y_i.

b)

Here we will use a tweak used in the lessons slides. If there are two possible global minimas,x_1, x_2 where $f(x_1) = f(x_2)$ and $x_1 > x_2$ in a convex set, we could write down the following statements :
- It exists a point t such that $f(t) > f(x_1) = f(x_2)$ which is in the set of points of the shortest path between x_1, x_2. This allows us to rewrite t with the convex equality and describe the point with respect to our 2 minimas:

$$t = \lambda x_1 + (1-\lambda)x_2, \lambda \in (0, 1]$$

and then reach the contradiction of:

$$f(t) = f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) = \lambda f(x_1) + (1-\lambda) f(x_1) = f(x_1)$$

Obviously $f(t)$ cannot be equal to $f(x_1)$, which is supposed to be our global minima. Hence, we prove that there are not 2 possible global minimas in this scenario, or in another way, our problem has an unique solution.

One extra requirement is that the set of points should be in general position.

c)

In the image we can see how - what are supposed to be the heaviest weights - there are weights with more rope. We can think of the optimal solution as the configuration of the center point closer to the heaviest ones. With this in mind, these points will have more rope and, therefore, will be closer to our base, where our $E_p = 0$.

So, the key here is to think that the optimal solution will minimize the distance between the heaviest points and our centrum and at the same time will minimze the potential energy because it is giving more rope to those points.