TP 5 : DESCENT METHODS

Introduction to descent methods and to their role in machine learning

Problem description

Machine learning requires to calibrate a pre-established function in order to match as best as possible with a a training set (inputs (x,…) & outputs v)

Ex : let $\mathbf{F}_p$ be the function to calibrate, with a set of parameters p to determine, and (x, …, v) the training set.

Questions

1. Pose this problem as an optimization problem
2. Is it constrained or unconstrained ? Why ?

Consider now a linear regression (F(x) = ax + b, where p=(a,b)) and let G(a,b) be the function to minimize

1. Is it differentiable or not ? Why ?
2. Write DGa and DGb (gradient of G in a and b respectively) formulas & implement them in python
3. Implement a gradient algorithm for this problem, on the training set given in the Excel file (trainingSet1.xlsx)
4. Compare your results with Excel linear regression

Gradient descent algorithms

1. Start from x_0
2. At Step k,
   - choose h^k such that $\nabla\varphi(x^k) \cdot h^k$ (HOW?)
   - choose small Step length $t^k > 0$ (HOW?), and update: $x_{k+1} = x_k + t^k h^k$
3. Compute $\nabla\varphi(x^{k+1})$
   - if ~ 0, STOP
   - Otherwise, go back to 2

$x = [x_1, \ldots, x_N]$

t : fixed (very small)

h_i = - gradient f in x_i

Stop when $f(x^{k+1}) - f(x^k) <$ tolerance

Main example of descent direction:

- $h^k = -\nabla\varphi(x^k)$: $\nabla\varphi(x^k) \cdot h^k \leq \nabla\varphi(x^k) \cdot h$ provided $|h| = |\nabla\varphi(x^k)|$
- Newton's method $h^k = -\left(\nabla^2\varphi(x^k)\right)^{-1}\nabla\varphi(x^k)$
- or $h^k = -Q^k\nabla\varphi(x^k)$, for some positive symmetric matrix Q^k