Optimization and Computational Linear Algebra for Data Science
Lecture 11: Gradient descent

Léo Miolane · leo.miolane@gmail.com

July 9, 2019

Warning: *This material is not meant to be lecture notes. It only gathers the main concepts and results from the lecture, without any additional explanation, motivation, examples, figures...*

1 Gradient descent

We aim at minimizing a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. Given an initial point $x^{(0)} \in \mathbb{R}^n$, the gradient descent algorithm follows the updates:

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)}), \tag{1}$$

where the step-size α_k remains to be determined. The step (1) is a very natural strategy to minimize f, since $-\nabla f(x)$ is the direction of steepest descent at x. Since $f(x+h) = f(x) + \langle \nabla f(x), h \rangle + o(\|h\|)$ we have

$$\begin{aligned} f(x^{(k+1)}) &= f(x^{(k)}) - \alpha_k \|\nabla f(x^{(k)})\|^2 + o(\alpha_k) \\ &< f(x^{(k)}) \end{aligned}$$

for α_k small enough (provided that $\nabla f(x^{(k)}) \neq 0$). Hence is the step-sizes α_k are chosen very small, the sequence $(f(x^{(k)}))_{k \geq 0}$ is decreasing!

2 Gradient descent for smooth convex functions

3 Gradient descent for smooth strongly convex functions

Algorithm 1 Gradient descent

Input: Graph Laplacian L, number of clusters k

1: Compute the first k eigenvectors $v_1, \dots, v_k$ of the Laplacian matrix L.
2: Associate to each node i the vector $x_i = (v_2(i), \dots, v_k(i)) \in \mathbb{R}^{k-1}$.
3: Cluster the points $x_1, \dots, x_n$ with (for instance) the k-means algorithm.

Further reading

See [2] for a very nice introduction to spectral clustering and [1] for lecture notes on spectral graph theory.

References

[1] Daniel Spielman. Spectral graph theory. *Lecture Notes, Yale University, http://www.cs.yale.edu/homes/spielman/561/2012/*, 2012.

[2] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.