Exercises Set 1

Paul Dubois

August 5, 2024

Abstract

Only the questions with a * are compulsory (but do all of them!).

1 Optimization

1.1 One dimension

Let's consider $f(x) = x^2 - x + 3$. We are interested in finding the minimum of this function; i.e. we want to find $x^* \in \mathbb{R}$ such that $f(x^*)$ is the smallest possible, mathematically:

$$\forall x \in \mathbb{R}, f(x) \geq f(x^*)$$

It turns out that since this function is quadratic, you can manually compute that $x^* = 0.5$. However, we will suppose that this is a complex function, for which we cannot find the minimum by hand. The techniques we will develop should be generalization to more complex functions, for which you won't be able to find the minimum "by hand".

Method 1 First, let's try to brute force our problem: compute $f(x)$ for $x = -1.75$, $x = -0.75$, $x = -0.25$, $x = 0.25$, $x = 0.75$, $x = 1.25$, $x = -1.75$, $x = 2.25$, $x = 2.75$, and $x = 3.25$.

Find the x value that gave you the smallest value for $f(x)$. This will be you first "guess" for x^*.

This technique is called "grid search".

List the positive and negative aspect of this technique.

Method 2 For this second approach, we will suppose that x^* is in $(-5, 5)$.

Define $a = -10$ and $b = 10$, and iterate the following process:

- if $f(a) > f(b)$, then let $a = \frac{a+b}{2}$

- if $f(a) < f(b)$, then let $b = \frac{a+b}{2}$

Manually do 6-7 steps of this process. Your final "guess" for x^* is $\frac{a+b}{2}$.

This technique is called "dichotomy".

List the positive and negative aspect of this technique.

Method 3 For this second approach, we will suppose that x^* is "close" to[1]
-0.5.
Define $x_0 = -0.5$, choose a 'learning rate' $\lambda = 0.8$ (say) and iterate the following
process:

- compute $f'(x_k)$

- let $x_{k+1} = x_k - \lambda * f'(x_k)$

(*) Manually do 5 steps of this process. Your final "guess" for x^* is the last x_k
calculated.
Repeat the process with $\lambda = 0.3$, $\lambda = 0.1$, $\lambda = 1$, $\lambda = 2$. (calculators /
computers are allowed!)
Give some conclusion on how to choose λ.
This technique is called "gradient descent".
List the positive and negative aspect of this technique.

1.2   Two dimensions

Let's consider $f(x,y) = (x+y-1)^2 + \frac{1}{5}(x-y)^2$. We are interested in finding the
minimum of this function; i.e. we want to find $(x^*, y^*) \in \mathbb{R}^2$ such that $f(x^*, y^*)$
is the smallest possible, mathematically:

$$\forall (x,y) \in \mathbb{R}^2, f(x,y) \geq f(x^*, y^*)$$

Again, it is possible to manually compute that the minimum is reached for
$(x^*, y^*) = (0.5, 0.5)$; however, we suppose that we do not know this.
    Generalize the three methods used in 1D to apply them in 2D.
    Explain how the computational cost evolve for each technique as the number
of dimensions increases.

2   Regression

(*) With justification, find **the** line of best fit of the following points:

- $P_1 = (0, 1)$

- $P_2 = (3, 8)$

- $P_4 = (5, 12)$

- $P_5 = (8, 17)$

---

[1]or at least "not too far" from