Exercises Set 1

Solutions

1 - Optimization

1.1 - One dimension

$f(x) = x^2 - x + 3$

$f'(x) = 2x - 1$

x	-1.75	-1.25	-0.75	-0.25	0.25	0.75	1.25	1.75	2.25
$f(x)$	7.8125	5.8125	4.3125	3.3125	2.8125	2.8125	3.3125	4.3125	5.8125

The two smallest values found are 2.8125, reached for $x = 0.25$ and $x = 0.75$. Hence, we can suppose the minimum reached for x in $(0.25, 0.75)$.

---

step	a	b	f(a)	f(b)
0	-10	10	113	93
1	0	10	3	93
2	0	5	3	23
3	0	2.5	3	6.75
4	0	1.25	3	3.3125
5	0	0.625	3	2.7656
6	0.3125	0.625	2.7852	2.7656
7	0.46875	0.625	—	—

(Table rounded to 10^{-4})

(2)

Step (k)	x_k	$f'(x_k)$	$f(x_k)$
0	-0,5	-2	3.75
1	1.1	1.1	3.11
2	0.22	-0.56	2.83
3	0.67	0.34	2.78
4	0.40	-0.20	2.76
5	0.56	–	–

$\lambda = 0.8$

(rounding to 10^{-2})

Step (k)	x_k	$f'(x_k)$	$f(x_k)$
0	-0.5	-2	3.75
1	0.1	-0.8	2.91
2	0.34	-0.32	2.76
3	0.44	-0.12	2.75
4	0.48	-0.04	2.75
5	0.49	–	–

$\lambda = 0.3$

(rounding to 10^{-2})

Step (k)	x_k	$f'(x_k)$	$f(x_k)$
0	-0.5	-2	3.75
1	-0.3	-1.6	3.39
2	-0.14	-1.28	3.16
3	-0.01	-1.02	3.01
4	0.09	-0.82	2.92
5	0.17	–	–

$\lambda = 0.1$

(rounding to 10^{-2})

Step (k)	x_k	$f'(x_k)$	$f(x_k)$
0	-0.5	-2	3.75
1	1.5	2	3.75
2	-0.5	-2	3.75
3	1.5	2	3.75
4	-0.5	-2	3.75
5	1.5	–	–

$\lambda = 1$

(rounding to 10^{-2})

step (k)	x_k	$f'(x_k)$	$f(x_k)$
0	-0.5	-2	3.75
1	3.5	6	11.75
2	-8.5	-18	83.75
3	27.5	54	731.75
4	-80.5	-162	6563.75
5	404.5	—	—

$\lambda = 2$

(rounding to 10^{-2})

Method	+	−
Grid Search (1)	• Easy to implement • No need for derivative	• Not very efficient • Needs bounds
Dichotomy (2)	• No need for derivative • Converges fast	• Needs bounds
Gradient Descent (3)	• Doesn't need bounds • Converges fast	• Needs an appropriate λ * • Need derivative

* λ is called the "learning rate"

1.2 - Two dimensions

$$f(x,y) = (x+y+1)^2 + \frac{1}{5}(x-y)^2$$

$$\frac{\partial f}{\partial x}(x,y) = 2(x+y+1) + \frac{2}{5}(x-y) = \frac{12}{5}x + \frac{8}{5}y + 2$$

$$\frac{\partial f}{\partial y}(x,y) = 2(x+y+1) - \frac{2}{5}(x-y) = \frac{8}{5}x + \frac{12}{5}y + 2$$

Grid search: calculate f on $S^2 = \{(x,y) \mid x \in S, y \in S\}$

with $S = \{-2.25, -1.75, -1.25, \ldots, 2.75\}$

find the pair $(x,y) \in S^2$ s.t. $f(x,y)$ is minimum.

Dichotomy: Start with $a = -10$, $b = 10$

$c = -10$, $d = 10$

if $f\left(a, \frac{c+d}{2}\right) < f\left(b, \frac{c+d}{2}\right)$: put $b = \frac{a+b}{2}$

if ______ > ______ : — $a = \frac{a+b}{2}$

if $f\left(\frac{a+b}{2}, c\right) < f\left(\frac{a+b}{2}, d\right)$: put $d = \frac{c+d}{2}$

if ______ > ______ : — $c = \frac{c+d}{2}$

Gradient Descent: Apply gradient descent on each component:

$(x_0, y_0) = (-0.5, -0.25)$ $\qquad \boxed{\lambda = 0.5}$

$$\frac{\partial f}{\partial x}(x_0,y_0) = \frac{6}{5} + \frac{2}{5} + 2 = \frac{18}{5}$$

$$\frac{\partial f}{\partial y}(x_0,y_0) = \frac{4}{5} + \frac{3}{5} + 2 = \frac{17}{5}$$

$$x_1 = x_0 - \lambda \cdot \frac{\partial f}{\partial x}(x_0,y_0) = \frac{1}{2} - \frac{1}{2} \cdot \frac{18}{5} = \frac{5-18}{10} = -\frac{13}{10}$$

$$y_1 = y_0 - \lambda \cdot \frac{\partial f}{\partial y}(x_0,y_0) = \frac{1}{4} - \frac{1}{2}\,\frac{17}{5} = \frac{10-34}{20} = -\frac{12}{10}$$

$(x_1, y_1) = (1.3, 1.2)$

[continue steps]

Method	Computational cost for n steps 1D	2D	KD
Grid Search (1)	n	n^2	n^k
Dichotomy (2)	n	2.n	k.n
Gradient Descent (3)	n	n or 2n (depending on how fast you compute partial derivatives)	n (if we compute partial derivatives all at once)

It is clear that for high dimensions, gradient descent is the most appropriate (assuming the function to optimize is differentiable).

Modern optimization algorithms are all variations of gradient descent; you can look at "momentum" or "Adam" if you are interested in modern techniques.

2 - Regression

Our model is $y = ax + b$.

We want to find the best a & b.

We have :

k	0	1	2	3
x_k	0	3	5	8
y_k	1	8	12	17

$\hat{y}_k = a x_k + b$

$e_k = \hat{y}_k - y_k$.

Let's build a function that measures the "performance" of the model given (a,b) :

$$F(a,b) = \sum_{k=0}^{3} (e_k)^2 = \sum_{k=0}^{3} (\hat{y}_k - y_k)^2 = \sum_{k=0}^{3} (ax+b-y_k)^2$$

$$= (a.0+b-1)^2 + (a.3+b-8)^2 + (a.5+b-12)^2 + (a.8+b-17)^2$$

The model perform the best with parameters (a^*, b^*) such that $F(a^*, b^*)$ is the minimum of F.

While we could use optimisation like in part 1 to find the minimum, there is an analytic solution in this case.

The minimum is reached when $\frac{\partial f}{\partial a}(a,b) = 0$ & $\frac{\partial f}{\partial b}(a,b) = 0$

(i.e. when $\nabla f = 0$)

$$\frac{\partial f}{\partial a}(a,b) = 2.(3a+b-8).3 + 2(5a+b-12).5 + 2(8a+b-17).8$$
$$= 2[9a+3b-24+25a+5b-60+64a+8b-136]$$
$$= 2[98a+16b-220]$$

$$\frac{\partial f}{\partial b}(a,b) = 2(b-1) + 2(3a+b-8) + 2(5a+b-12) + 2(8a+b-17)$$
$$= 2[16a+4b-38]$$

So we want $\begin{cases} 98a + 16b = 220 \\ 16a + 4b = 38 \end{cases}$

$\Leftrightarrow \begin{cases} 34a = 68 & (L_1 - 4.L_2) \\ 4b = 38 - 16a & (L_2 \text{ rearranged}) \end{cases}$

$\Leftrightarrow \begin{cases} a = 2 \\ b = 1.5 \end{cases}$

[There is only one minimum since our function f is convex.]

The best linear model is acheived for $a=2$, $b=1.5$.