Optimisation

(Course: Introduction to Data Science)

Tirtharaj Dash

BITS Pilani, K.K. Birla Goa Campus

tirtharaj@goa.bits-pilani.ac.in

October 30, 2020

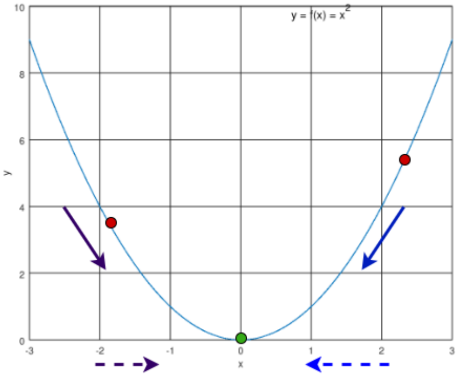
Optimisation I

Let's look at minimisation problems for functions that are continuous and differentiable.

- If the derivative of the function is positive, the function is increasing.
  - Don't move in that direction, because you'll be moving away from a minimum.
- If the derivative of the function is negative, the function is decreasing.
  - Keep going, since you're getting closer to a minimum.

Optimisation II

Let $f(x) = x^2$. The function looks like this:



The arrows show movement of next functional value, and the dotted arrows show the corresponding direction of movement of x.

Optimisation III

Here is a very simple gradient descent procedure:

1. Initialize x to some value
2. **while** stopping criterion is not met
    1. Calculate the gradient of the function, $\nabla_x f$
    2. $x := x - \eta \nabla_x f$
3. **return** x

Notice step 2.2. above: x will move right, if $\nabla_x f$ is negative, and it will move left, if $\nabla_x f$ is positive.

Notation: I use $\nabla_x f$ to denote $\frac{\partial f}{\partial x}$.

1. Using gradient descent, obtain the value of x that minimizes $f(x) = (x-2)^2 - 5$. Starting value of $x = 3$ and $\eta = 1$.

   Answer. Derivative of f w.r.t. x: $\nabla f = 2(x-2)$

   - $x = 3$: $\nabla f|_{x=3} = 2$; $x = 3 - 2 = 1$; $f(1) = -4$
   - $x = 1$: $\nabla f|_{x=1} = -2$; $x = 1 - (-2) = 3$; $f(3) = -4$.
   - ... gets repeated.

2. Solve the same question with same starting point, but with $\eta = 0.5$.

Answer. Derivative of f w.r.t. x: $\nabla f = 2(x-2)$

- $x = 3$: $\nabla f|_{x=3} = 2$; $x = 3 - 0.5 \times 2 = 2$; $f(2) = -5$
- $x = 2$: $\nabla f|_{x=2} = 0$; $x = 2 - 0.5 \times 0 = 2$; $f(2) = -5$.
- $x = 2$: $\nabla f|_{x=2} = 0$; $x = 2 - 0.5 \times 0 = 2$; $f(2) = -5$.
- Value of f doesn't change further. So, stopping criterion met. Return $x = 2$. This is same as the exact solution i.e. Find root of $\nabla f = 0$.

Optimisation VI

Gradient descent is guaranteed to eventually find a local minimum if:

- the learning rate is set appropriately (sometimes, using adaptive learning rate); $\eta \in [0.0001, 1]$.
- a finite local minimum exists (i.e. the function doesn't keep decreasing forever).

Optimisation VII

Various stopping criteria for gradient descent:

- Stop when the norm of the gradient is below some threshold, θ

$$||\nabla f|| < \theta$$

  This is checking the distant the gradient is from the origin, 0.

- Maximum number of iterations is reached.

It is straightforward to extend the gradient descent procedure to scalar functions with multiple variables.

3. Let $f(x_1, x_2) = 3x_1^2 - 2x_1x_2 + x_2^2 - 5$. Initial values $x_1 = 1$, $x_2 = 1$. Fix $\eta = 1$.

   Answer. Present value of f: $f(1,1) = 3 - 2 + 1 - 5 = -3$. The partial derivatives are:

$$\nabla_{x_1} f = 6x_1 - 2x_2$$
$$\nabla_{x_2} f = 2x_2 - 2x_1$$

Optimisation IX

Update the present $x_{1,2}$:

$$\begin{aligned} x_1 &= x_1 - \eta \nabla_{x_1} f \\ &= 1 - (6 - 2) = -3 \\ x_2 &= x_2 - \eta \nabla_{x_2} f \\ &= 1 - (2 - 2) = 1 \end{aligned}$$

New value of f: $f(-3, 1) = 29$. Update the present $x_{1,2}$ using gradients:

$$\begin{aligned} x_1 &= x_1 - \eta \nabla_{x_1} f \\ &= -3 - (-18 - 2) = 17 \\ x_2 &= x_2 - \eta \nabla_{x_2} f \\ &= 1 - (-6 - 2) = 9 \end{aligned}$$

New value of f: $f(17, 9) = 637$.

4. Solve the above question with $\eta = 0.1$.

Answer. Update the present $x_{1,2}$:

$$x_1 = 1 - 0.1(6 - 2) = 0.6$$
$$x_2 = 1 - 0.1(2 - 2) = 1$$

New value of f: $f(0.6, 1) = -4.12$. Update the present $x_{1,2}$ using gradients:

$$x_1 = 0.6 - 0.1(3.6 - 2) = 0.44$$
$$x_2 = 1 - 0.1(2 - 1.2) = 0.92$$

New value of f: $f(0.44, 0.92) = -4.38$.

This is function surface and its contour: