Artificial Intelligence Lab 6: Gradient Descent

Let $L\colon \mathbb{R} \to \mathbb{R}$ be a function, and let:

(1) $L_* = \min_{w\in\mathbb{R}} L(w)$ be its **minimum** value

(2) $w_* = \arg\min_{w\in\mathbb{R}} L(w)$ be the **minimiser**, i.e., the argument for which the minimum value is attained (note that $L_* = L(w_*)$).

The gradient descent algorithm to find w_* of a function $L(w)$ is given as follows:

$$w_{t+1} = w_t - \alpha \frac{dL(w)}{dw}|_{w=w_t}, \tag{1}$$

where

(i) $\alpha > 0$ is a step-size parameter

(ii) $\frac{dL(w)}{dw}|_{w=w_t}$ is the derivative of $L(w)$ with respect to w evaluated at $w = w_t$.

Let $L(w) = \frac{a}{2}w^2 + bw + c$.

Q1) [20 Marks] Plot the function $L(w)$ (in red) $L_* = L(w_*)$ (blue dot) and w_* (green star), for the following cases:

(1) $a = 1, b = 0, c = 0$.

(2) $a = 0.1, b = 0, c = 0$

(3) $a = 10, b = 0, c = 0$

(4) $a = 1, b = 0, c = 10$

(5) $a = 1, b = 1, c = 1$

(6) $a = 0.1, b = -1, c = -1$

Calculate L_* and w_* with pen and paper. Choose appropriate axis range so that L_* and w_* can be displayed in a proper manner.

Q2) [30 Marks] Run the gradient descent algorithm for $t = 1, \ldots, T$ iterations (try various values for T such as $10, 100, 1000$). For each iteration display

(1) The complete function $L(w)$ (for w in an appropriate range used in [Q1]), and $L(w_t)$ in one plot

(2) $w_t - w_*$ in a different plot.

Run the gradient descent algorithm for various values of α such that : (i) the iterates w_t are on the same side of w_* and converge to w_*, (ii) the iterates w_t oscillate on both sides of w_* and converge to w_*, (iii) the iterates w_t oscillate on both sides of w_*, but diverge to ∞. Please derive these cases with pen and paper first and then proceed to code.

Q3) [25 Marks] Create your own function with multiple local minima, and compare (i) gradient descent versus (ii) gradient descent with momentum.

Q4) [25 Marks] Gradient Descent in 2D: Let $w \in \mathbb{R}^2$. Consider the functions $f_1(w) = \frac{1}{2}w(1)^2 + \frac{1}{2}w(2)^2$, $f_2(w) = \frac{10}{2}w(1)^2 + \frac{1}{2}w(2)^2$, $f_3(w) = \frac{1}{2}w(1)^2 + \frac{10}{2}w(2)^2$, $f_4(w) = \frac{1}{2}w(1)^2 + \frac{1}{2}w(2)^2 + 5w(1) - 3w(2) - 2$. For the functions $f_i, i = 1, \ldots, 4$

a) Show the negative gradient directions and contour plots.

b) Perform gradient descent to find the minima and show the trajectories of the gradient descent algorithm. Use different step-sizes to demonstrate (i) no oscillation, (ii) oscillation in one coordinate (iii) one-sided, oscillatory and divergent modes of convergence behavior.