Lecture 9

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Convex Functions

Convex Functions are defined mathematically, as follows-

$$\alpha L(w_1) + (1-\alpha)L(w_2) \geq L(\alpha w_1 + (1-\alpha)w_2),$$

where, α is between 0 and 1.
In L2 regularization, we work under the assumption that $L(w)$ is convex.

Adding two convex functions, with a positive multiplier for each, the resulting function is also a convex function. (`Jensen's Inequality`).

Gradient Descent Algorithm

A very important technique in all of modern AI-ML tasks.

Basic intuition - Even if we can't calculate the exact minimum, we can still get closer to the minimum iteratively in small steps.

Randomly initialize w
Until we reach terminal condition.

$w_{\text{new}} \leftarrow w_{\text{old}} - \eta \nabla L_w$
$w_{\text{old}} \leftarrow w_{\text{new}}$

η is the learning rate (or step size) and needs to be tuned properly in real-world systems.

Perform this in all components of w (dimension of w).

Now, the terminal conditions could be one (or a combination) of the following -

1. `max_iter` has been reached
2. $L(w_o) - L(w_n) < \epsilon$
3. $(w_o - w_n)^T(w_o - w_n) < \delta$

Since, from the last lecture Lecture 8, we has the expressions for E_D and E_W, hence we can compute the gradient for those expressions and we would get the update-iterate step.