Subjective Question

- 2018101116

Gradient Descent optimizes an objective or a loss function with the help of an iterative step.

where

$$w^{k+1} \leftarrow w^k - \eta \nabla J$$

new Sol^n

If update was more generic : $w^{k+1} \leftarrow w^k + \delta$

$\boxed{\delta = -\eta \nabla J}$ for GD

i.e $\quad w^{k+1} - w^k = \delta$

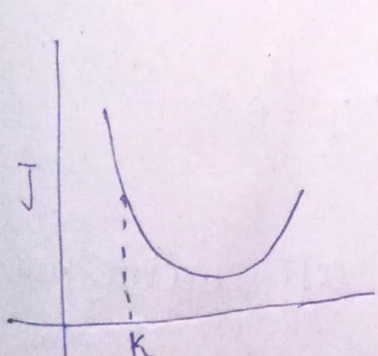
Using Taylor's series ; which is given by

$$f(y) = f(x) + (y-x) f'(x) + \frac{(y-x)^2}{2} f''(x) + \cdots$$

Assuming y is a vector we can write

$$f(y) = f(x) + [y-x]^T \nabla f + \frac{1}{2} [y-x]^T M [y-x]$$

Hassian matrix



Given any time 'k', we would like to complete the objective f^n J at $k+1$ nth instruct;

This is given by

$$J(w^{k+1}) \approx J(w^k) + [w^{k+1} - w^k]^T \nabla J(w^k) + \frac{1}{2} [w^{k+1} - w^k] M [w^{k+1} - w^k]$$

or

$$\boxed{J(w^{k+1}) = J + s^T \nabla J + \frac{1}{2} s^T H s}$$

Now to check if the error reduces, let us start with a further truncated expression

$$J(w^{k+1}) = J(w^k) + s \nabla J$$

$$J(w^{k+1}) = J(w^k) - \eta \nabla J^T \nabla J$$

$$J(w^{k+1}) = J(w^k) - \eta \, [(+ve) \text{ term}]$$

Thus, this new objective is smaller than the previous objective.

Improvement of objective further depends on η

$$J(w^{k+1}) = J(w^k) + s^T \nabla J + \tfrac{1}{2} s^T H s$$

$$= J - \eta \nabla J^T \nabla J + \frac{\eta^2}{2} \nabla J^T H \nabla J \quad [\text{quadratic in } \eta.$$

To optimize this, we differentiate w.r.t η & equate to zero

This gives $-\|\nabla J\|^2 + \eta \nabla J^T H \nabla J = 0$

or $\boxed{\eta = \dfrac{\|\nabla J\|^2}{\nabla J^T H \nabla J}}$ This is the best update rule.

A function $f(x)$ is convex if $f''(x) \geq 0$ for all x.
i.e $f'(x)$ is an increasing function of x.

The minimum is attained when $f(x) = 0$ since $f(x)$ keeps increasing to the left & right of that.

Thus there exists a global minimum which is unique.

For multivariate functions $\nabla^2 f(x)$ or the second derivate is the Hessian matrix H.

We have already established that $\nabla J^T \nabla J$ is positive and if H is positive semidefinite the convex.

Given a function $f(x) = x^T A x$, where A is positive semidefinite, Hessian of $f(x)$ is A

$\therefore$ H in $\nabla J^T H \nabla J$ is positive semidefinite.

Scanned with CamScanner