Note Title 8/5/2021

$\min_x f(x)$

Gradient descent

update rule : $x_{t+1} = x_t + \eta_t \left(-f'(x_t)\right)$

↳ what is so special about $-f'(x)$?

Taylor's series : $f(x+\eta d) = f(x) + \eta d f'(x) + \frac{\eta^2 d^2}{2} f''(x) + \cdots$

$\bar{x} = x + \eta d$

evaluations are all at x

Local information gives global information

Note Title 8/5/2021

$$f(x+\eta d) = f(x) + \eta d f'(x) + \underbrace{f''(x)\frac{\eta^2 d^2}{2} + \cdots}_{\text{higher order terms}}$$

Small positive step-size

'direction'

For small enough η

$$f(x+\eta d) \approx f(x) + \eta d f'(x)$$

$$\Rightarrow \boxed{\underbrace{f(x+\eta d)}_{\substack{\text{Function evaluation}\\ \text{at updated point}\\ \text{along "direction" } d}} - \underbrace{f(x)}_{\substack{\text{Function evaluation}\\ \text{at the current}\\ \text{point}}}} \approx \underline{\eta d f'(x)}$$

Want to choose a direction d s.t

$$\boxed{f(x+\eta d) - f(x) < 0}$$

$\Rightarrow$ Want 'd' s.t

$$\eta d f'(x) < 0$$

Note Title 8/5/2021

Want $\eta \, d \, f'(x) < 0$ $\implies$ Want d s.t $\boxed{d \, f'(x) < 0}$

$\uparrow$

Small +ve constant

For the choice of $\boxed{d = -f'(x)}$

$$d \, f'(x) = -\left(f'(x)\right)^2 < 0$$

Higher dimensions

$$f(x_1, x_2) = x_1^2 + 4x_2 + 8x_2^2$$

Derivative $\iff$ Gradient

vector of partial derivatives.

$$\nabla f\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{x_1 = a} \\ \left.\frac{\partial f}{\partial x_2}\right|_{x_2 = b} \end{bmatrix}$$