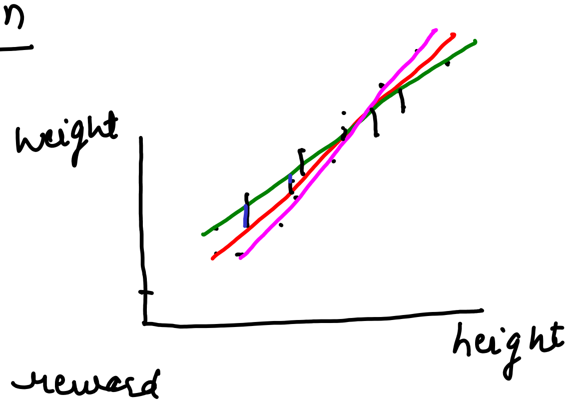
Optimization

Find the 'best' line.



$y = \check{m}x + \check{c}$

—

Prediction → least loss

Reinforcement learning → maximum reward

General form of optimization Problem:

$$\boxed{\min_{x \in \mathbb{R}^d} \; f(x)} \rightarrow \text{objective } f^n \qquad \max_{x} \; f(x)$$

$$\boxed{\max_{x \in \mathbb{R}^d} \; -f(x)}$$

$g_i(x) \leq 0 \quad \forall i = 1, \ldots k$ inequality constraint

$h_j(x) = 0 \quad \forall i, \; 1, \ldots l$ equality constraint

---

$$\min_{x} \; x^3 + 2x^2 + 3$$

$$x + 2 \leq 4$$

$$2x + 1 = 3$$

Types of optimization Problems:

1. Constrained ✓
2. Unconstrained ✓

—— × ——

How to solve an optimization Problem? (unconstrained)

$$\min_{x \in \mathbb{R}} \; (x-5)^2 \qquad\qquad x = 5$$

$$f(x) = (x-5)^2$$

$$f'(x) = 2(x-5) = 0$$

$$x - 5 = 0$$

$$x = 5$$

$$\min_{x \in \mathbb{R}} 3x^6 + 2x^5 + 3x^3 + 5x^2 + 2$$

$$f'(x) = \underline{18x^5 + 10x^4 + 9x^2 + 10x} = 0$$

$f'(x) = 0$

We need a general method. **Gradient Descent**

A systematic procedure to optimize a f^n:

1. Start with $x_0 \in \mathbb{R}$ [arbitrary choice]
2. For $t = 1, \ldots, T$:
   update x
   $$\boxed{x_{t+1} = x_t + d}$$ → direction

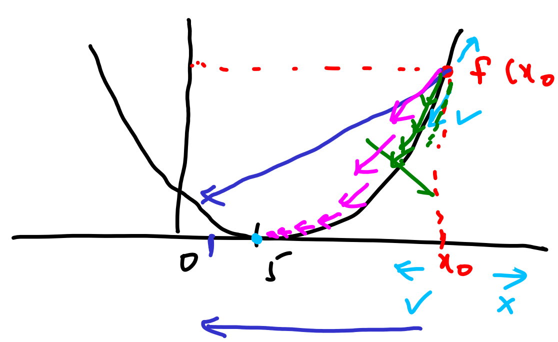
end

What is a good 'd'?

'd' is a direction that hopefully takes us closer to the minima of the f^n.

$(x-5)^2$

How much to move in direction 'd'?



$\eta = 0.01$

$\eta = 0.0001$

1000

'η' → step size

$\vec{W}_{t+1} = \vec{W}_t + \eta \vec{d}$

L.R.

$F(x_t)$

$$\boxed{x_{t+1} = x_t + \eta d}$$

New old

$\hat{y} = w_0 x_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_m x_m$

$f(x_{t+1}) < f(x_t)$ ✓

$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = ?$ → Min loss f^n

η = small +ve value

0.1
0.01

Loss

w_0

$x_t \rightarrow x_{t+1}$

max_iter
= 10000

We need to know something about the future value of f at x_{t+1} to be able to prudently decide on the value of d.

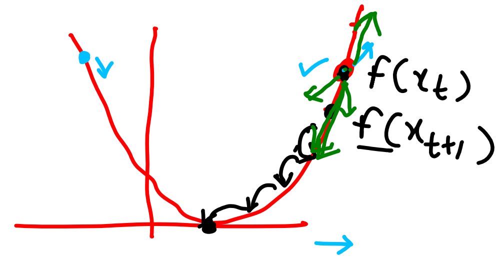
η ① constant
② adaptive

$f(x_t)$ $f(x_{t+1})$ Taylor's series

local info $\rightarrow$ Global

x

$$\left.\begin{array}{l} f(x) \\ f'(x) \\ f''(x) \\ \vdots \\ f''''(x) \end{array}\right\} \Rightarrow f(x+\Delta x)$$



$\eta =$

$$f(x+\Delta x) = f(x) + (\Delta x)\frac{f'(x)}{1!} + (\Delta x)^2 \frac{f''(x)}{2!} + \dots$$

$$\boxed{f(x_t+\eta d) = f(x_t) + (\eta d)\frac{f'(x_t)}{1!}} + \underbrace{(\eta d)^2 \frac{f''(x_t)}{2!}}_{\text{can ignore for small } \eta} + \dots$$

$$f(x_t+\eta d) - f(x_t) = \eta d f'(x_t)$$

$\Downarrow$ $f(x_{t+1})$ f^n value at the new point we wish to move to

$\Downarrow$ f^n value at the current point

$d = -f'(x_t)$ ✓

$d = -\dfrac{1}{f'(x_t)}$

$$f(x_t+\eta d) - f(x_t) < 0$$

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

$d = ?$

η = small +ve value

$d \, f'(x_t) < 0$

$d = ?$

$f'(x_t)$ depends on f and x_t

if $f'(x_t)$ is $-$ve, d should be +ve

if $f'(x_t)$ is +ve, d should be $-$ve

$d = ?$

$$\boxed{d = -f'(x_t)}$$

Update Rule:

$$x_{t+1} = x_t + \eta d$$

$$\boxed{x_{t+1} = x_t - \eta f'(x_t)}$$

Loss fn

MSE

$x \in \mathbb{R}$

$x = \begin{bmatrix} \mathbb{R} \\ \mathbb{R} \\ \mathbb{R} \end{bmatrix}^{d}$ eg.

HPT

Higher Dimensions:

Gradient is a vector

of Partial derivatives.

$$\vec{x}_{t+1} = \vec{x}_t + \eta \vec{d}$$

$$f(\vec{x}_t + \eta \vec{d}) = f(\vec{x}_t) + \eta \vec{d}^T \cdot \vec{\nabla f}(\vec{x}_t) + \ldots$$

$$\underbrace{f(\vec{x}_t + \eta \vec{d}) - f(\vec{x}_t)} = \eta \vec{d}^T \nabla f(\vec{x}_t)$$

$$\eta \vec{d}^T \nabla f(\vec{x}_t) < 0$$

$$\vec{d}^T \nabla f(\vec{x}_t) < 0$$

$$\boxed{d = -\nabla f(\vec{x}_t)}$$

$f(x, y) = 2x + 3y + 2$

$\frac{\partial f}{\partial x} = 2$ $\qquad$ $\frac{\partial f}{\partial y} = 3$

$\nabla f(x, y) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

update rule:

$$\boxed{\vec{x}_{t+1} = \vec{x}_t - \eta \nabla f(\vec{x}_t)}$$

$A * B$

$m \times \underline{n} \quad \underline{m} \times m$

m

$u_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$

$u_1 \cdot u_2 = 1*2 + 2*3 + 3*4$

$= 2 + 6 + 12 = \underline{20}$

$u_1^T u_2 = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$

$1\times3 \quad 3\times1$

1×1

$= \underline{20}$

$\max f(x)$

$\min -f(x)$

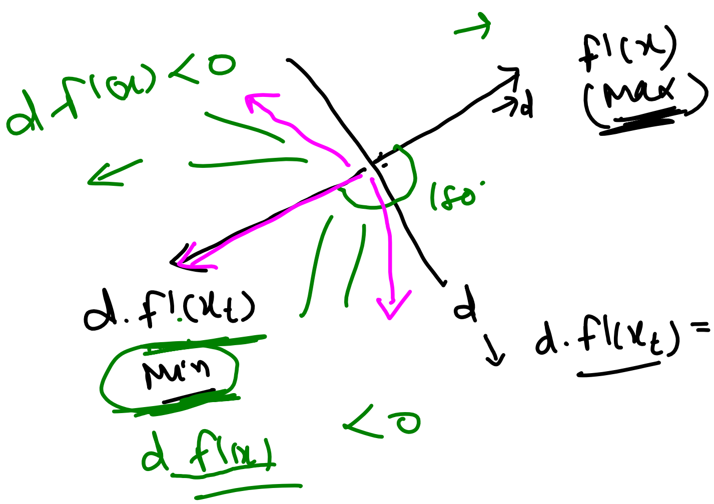
$a \cdot b$

$= a\, b \cos\theta$

$d^T f'(x_t) < 0$

$d \cdot f'(x_t) < 0$

$d = ?$



$\overset{?}{d}\ f'(x_t) \cos\theta$

$\cos 90 = 0$

$\cos 0 = 1$

$\cos 180° = -1$

$f(x)$ ✓

x_0

$\vec{x}_{t+1} = \vec{x}_t - \eta \nabla f(\vec{x}_t)$

$f(x) = x^3 - x^2 - x + 5$

$x_0 = 0.75$

$\eta = 0.25$

① Traditional

Code:

10 iterations

② GD: I_1: $x_1 = x_0 - \eta f'(x_0)$

I_2: $x_2 = x_1 - \eta f'(x_1)$

$(x_0, y_0) =$

η

$f(x,y) = 2xy + 2x - x^2 - 2y^2$

$\cancel{f'(x,y)}$ $\dfrac{\partial f}{\partial x} =$ $\dfrac{\partial f}{\partial y} =$ $\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix}$

I_1: $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \eta \begin{bmatrix} \\ \\ \end{bmatrix}_{(x_0, y_0)}$

I_2:

$$\underline{\underline{\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}}} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \eta \begin{bmatrix} \\ \\ \end{bmatrix}_{(x_1, y_1)}$$