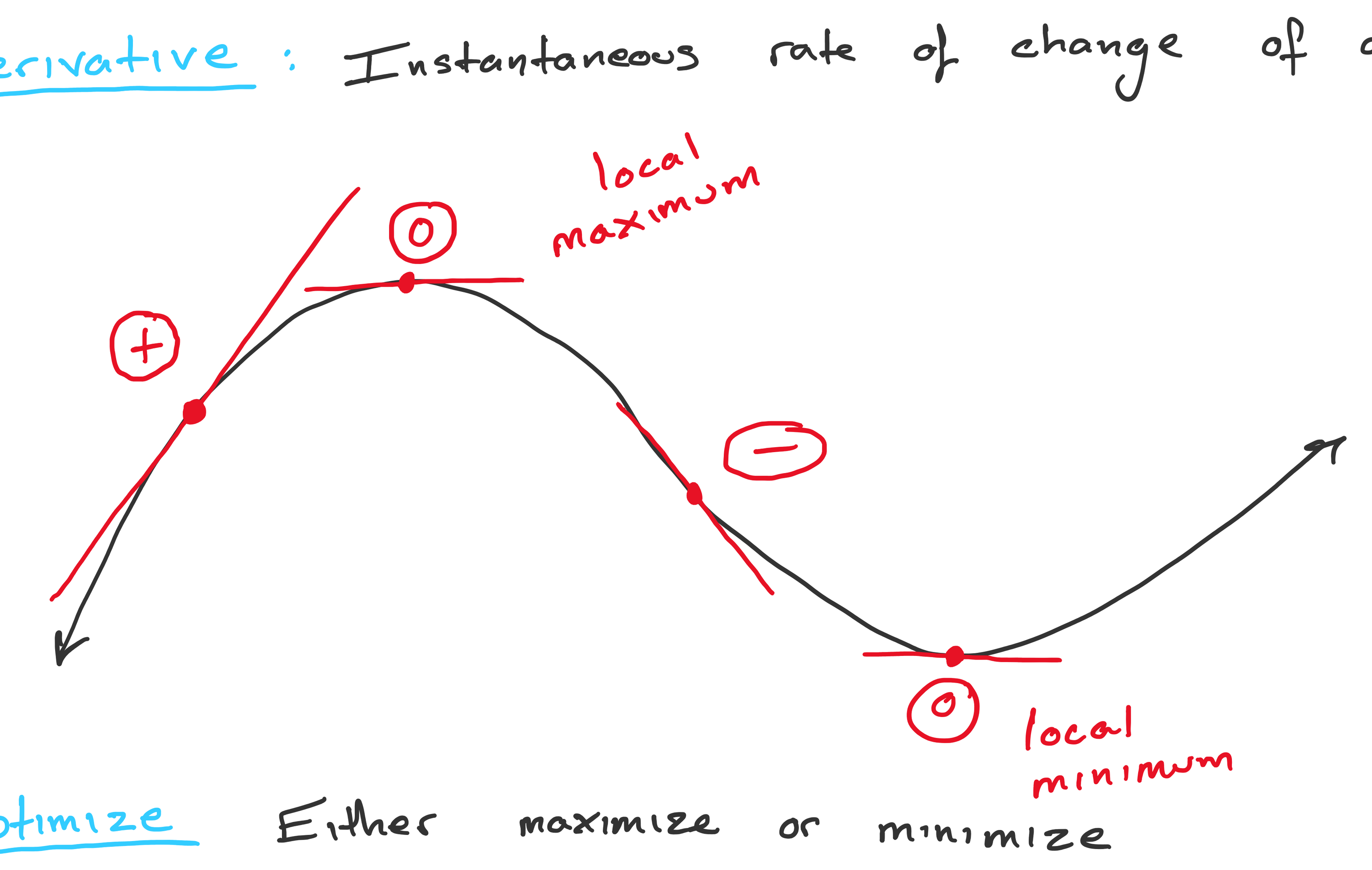
Grad Desc - EC

Thursday, October 17, 2019 9:55 AM

Defs

Derivative: Instantaneous rate of change of a fn



Optimize: Either maximize or minimize

Loss Function: A function that tells us the discrepancy btwn y and $\hat{y}$.

$$\text{MSE} = \frac{1}{n}\sum (y_i - \hat{y}_i)^2$$

$$\text{LASSO loss} = \text{MSE} + \lambda \|\beta\|_1$$

Classification error = 1 − acc ← NOT DIFFERENTIABLE!

Log Loss (Binary Crossentropy) $= \frac{1}{n}\sum\left[-y_i \log \hat{p}_i - (1-y_i)\log(1-\hat{p}_i)\right]$

SVM Hinge loss $= \frac{1}{n}\sum \max\{1 - y_i f(x_i), 0\} + \frac{c}{2}\|w\|^2$

Old way to optimize:

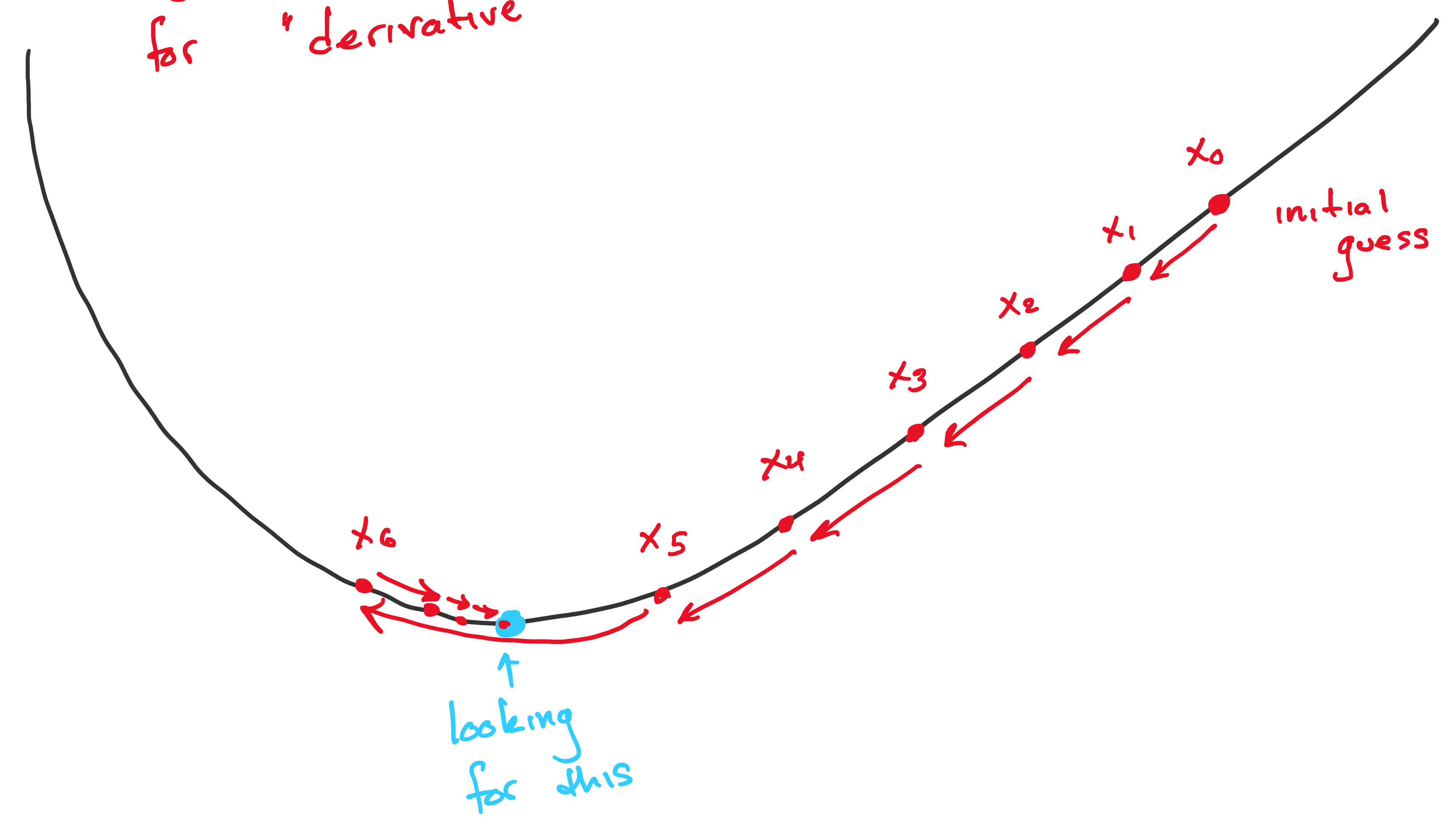
minimize f: $f'(x) = 0 \Rightarrow$ solve for x

Convex: Any U-shaped function. $f''(x) \geq 0$ for all x

Convex: Non convex

Gradient Descent: → go downhill

engineer-speak for "derivative"



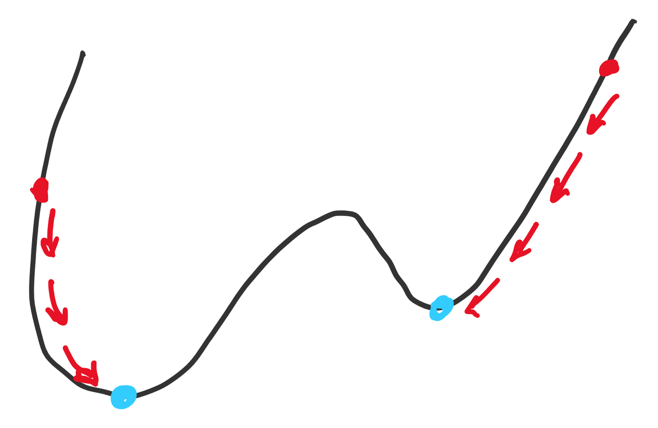
For step $k+1$:

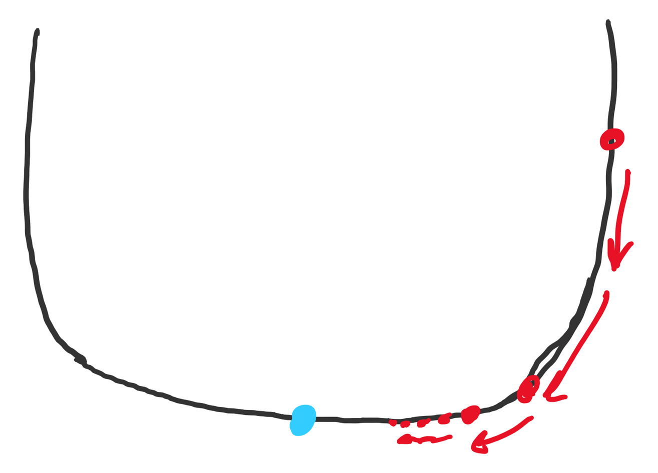
$$\boxed{x_{k+1} \leftarrow x_k - \alpha f'(x_k)}$$

new step; current step; move in opp dir of deriv; step size; gradient at current step

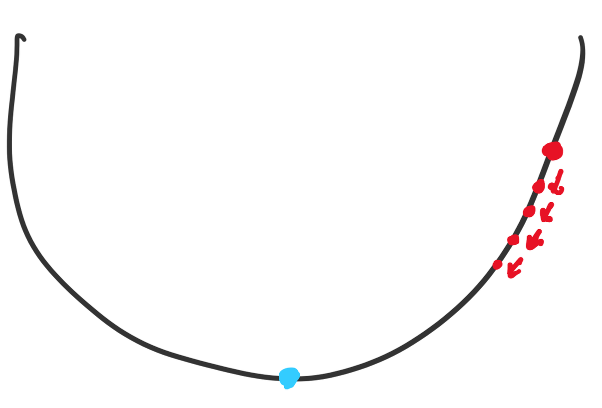
What can go wrong?

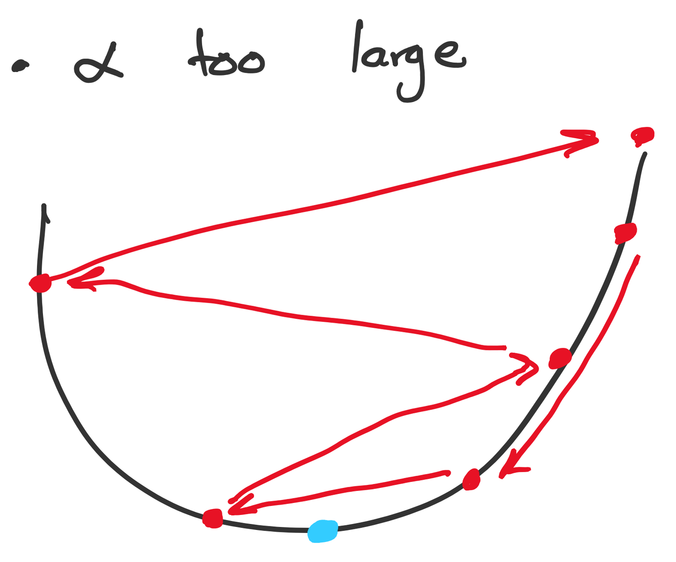
- Local minima
- Function too flat





- α too small
- α too large





How to pick α:

- Lazy way: guess and check
- $\alpha = \frac{c}{k}$
- Newton's method:

$$x_{k+1} \leftarrow x_k - \frac{f'(x_k)}{f''(x_k)} \quad \text{Sometimes hard or impossible to find}$$

$$x_{k+1} \leftarrow x_k - \underbrace{[\nabla^2 f(x_k)]^{-1}}_{\text{matrix}} \underbrace{\nabla f(x_k)}_{\text{vec}}$$

- Use fancier algorithms
  - RMSProp
  - Adam

minimize $f(x) = -\frac{\log x}{1+x}$

$$f'(x) = -\frac{1/x + 1 - \log x}{(1+x)^2} = 0$$

$$\frac{1}{x} + 1 - \log x = 0$$

$$1/x + 1 = \log x$$

$$e^{1/x + 1} = x$$