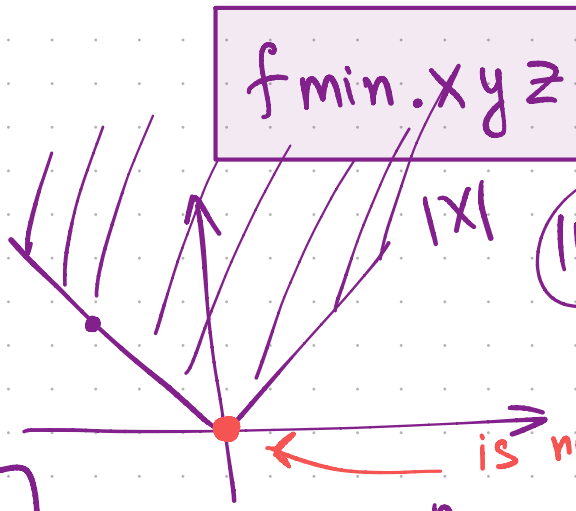
Optimization Methods.

msai 22. fmin.xyz.

materials will be placed here.

Daniil Merkulov



① $f(x) = |x|$ — 100%

is not differentiable

② $A \in \mathbb{S}^n_{++}$ — convex — 84%

$\forall x \in \mathbb{R}^n$:

$x^T A x > 0$

$A, B \in \mathbb{S}^n_{++}$

$C = \theta A + (1-\theta) B \in \mathbb{S}^n_{++}$

$x^T C x = x^T \theta A x + x^T (1-\theta) B x$

$\theta x^T A x + (1-\theta) x^T B x > 0$

$\geq 0 \quad > 0 \quad \geq 0 \quad > 0$

③ Majority do l_1 regularization

l_1

$Ax = b \Rightarrow$ you want to find the solution

$$x \begin{pmatrix} 6.873 \\ 0.0001 \\ -0.000007 \\ \vdots \\ 1.15 \\ 0.1 \\ -1.2 \end{pmatrix}$$

$$\|x\|_1 + \|Ax - b\|_2^2 \to \min_{x \in \mathbb{R}^n}$$

$$x \begin{pmatrix} 7 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ -1 \end{pmatrix}$$

$N = 10\,000$ train samples

$b = 100$

SGD steps per epoch

$\frac{10000}{100} = 100$ per epoch

10 epochs

Linear Programming

30 years PROGRESS

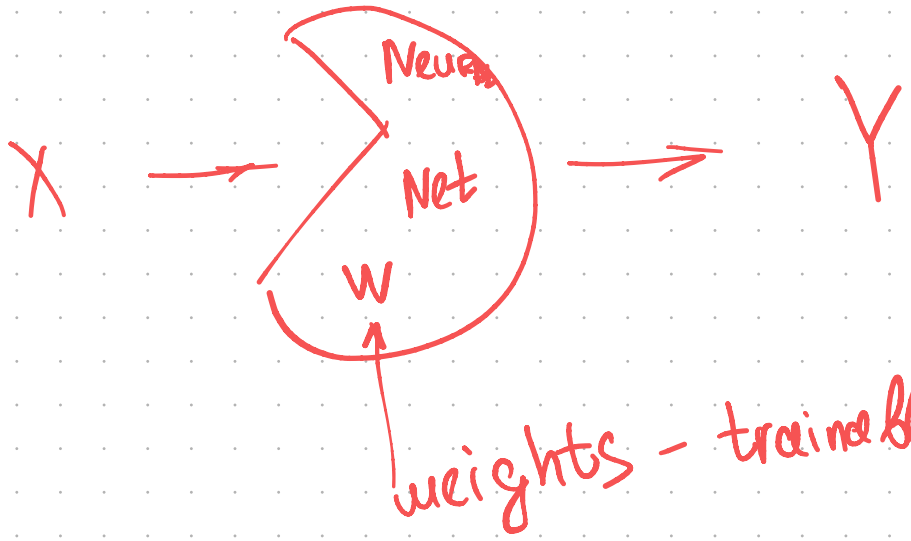
Computational Resources PROGRESS

Algorithmic PROGRESS

1. Solving LP with modern machines but with an old algor. soft.

2. Solving LP with old machines but with the modern algorith

Algorithmic PROGRESS >> Res. PROGRESS.



$$f(x) = W \cdot X + b \approx Y$$

$$f(x) = \sigma(W \cdot X + b) \approx Y$$

↑ non-linearity

PROBLEM NON-CONVEX.

Neural Network Optimization !!

Lipschitz constant

$$f(x) = Ax - b$$

$$f: \mathbb{R}^n \to \mathbb{R}^m$$

$$\nabla f =$$

$$\|\nabla f\| \leq \|A\|$$

$$|f(x) - f(y)| \leq L \cdot \|x - y\|$$

$$| : \|x - y\|$$

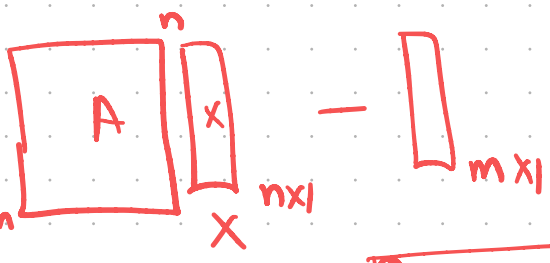
$$L = \|A\|$$

$f'(x) \approx$

$$\frac{f(x) - f(y)}{\|x - y\|} \leq L$$

$\frac{\partial f_i}{\partial x_j}$ — by defin → using df notation

$f(x) = Ax - b$



MATRIX CALCULUS??

BRIEF RECAP

SCOPE

2 ~1.5h.

1 lect. — lectures — PRE RECORDED VIDEOS with brief th. intro to the topic

1 sem. — seminars — ONLINE with interactive ex.

7 topics — 7 pairs — lecture seminar

1. Automatic different. Examples. Autograd, Breif Recap ~~of the~~ matrix calculus
2. Markowitz Portfolio Theory. LP
3. Zero order Methods. Global optimization

ML hyperparameters tuning

④ the Concept of Barrier methods

LASSO, Ridge Regression.

⑤ Quasi Newton Methods (BFGS, L-BFGS)

⑥ Stochastic gradient algorithms. Batch

⑦ NEURAL NETWORKS Epochs, Adam.

Loss surface

landscape.

# of PROBLEMS → seminar →

→ HW

???

SDP,

Duality

???

Integer programming

MIP.