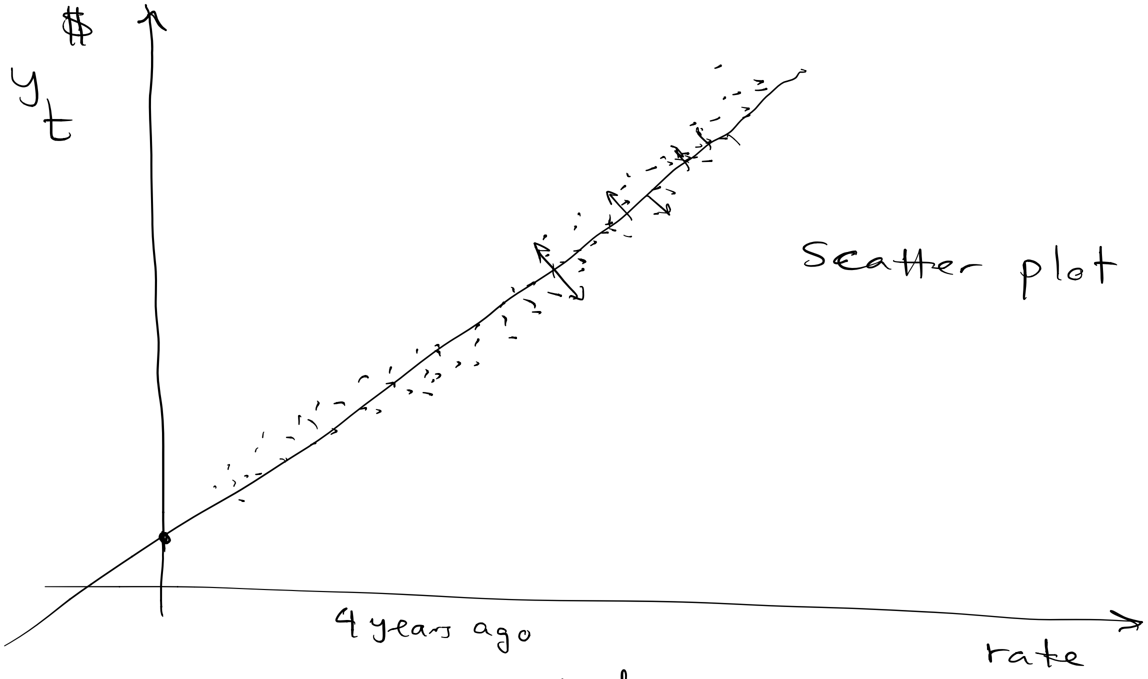


today

(r_t, p_t) $t=1,\dots,T$ $\qquad x_t$

~~y~~ $$y_t = \boxed{a_0} + b_1 x_t^{(1)} + \overset{\text{white noise}}{\underset{1}{\boxed{\sigma}}\,\varepsilon_t}$$

Goal: To find a_0 & b_1

$$y_t = \underbrace{\underbrace{\underbrace{\underbrace{\boxed{a_0} + \boxed{b_1}} x_t^{(1)} + \boxed{b_2}} x_t^{(2)} + \boxed{b_3}} x_t^{(3)} + \cdots}$$

$$\boxed{y_t = a_0 + b_1 x_t^{(1)} + \varepsilon_t}$$

1. linear regression (regress)

a_0 & b_1 & Statistics on R^2

$R^2 \leq 1$

2. Simplex

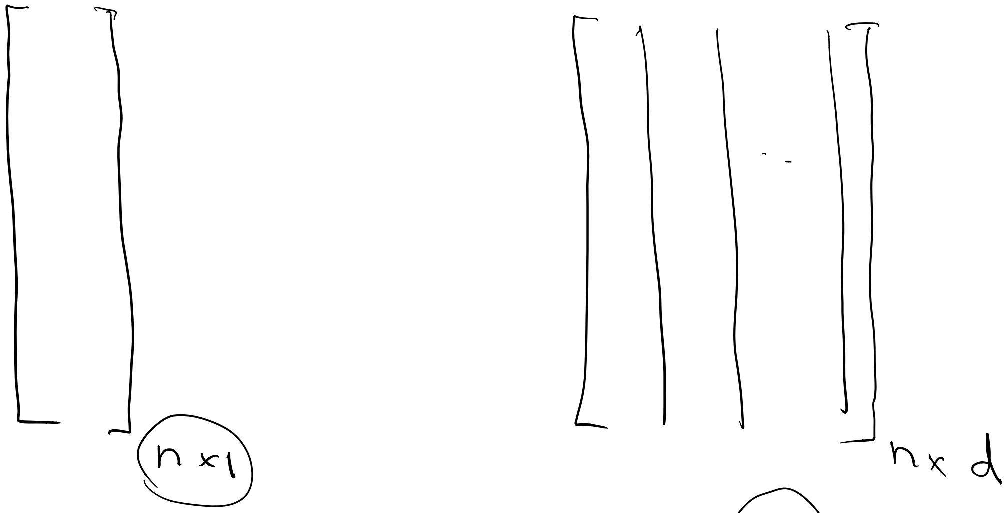
loss function

3. Gradient Descent (vanilla)

- SGD & Batch

Y ↑ observation

X ↑ dependence



$$\boxed{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad n \times 1$$

$$\boxed{X} = \begin{array}{cc} \left[\begin{array}{cc} 1 & x_1^{(1)} \\ \vdots & x_2^{(1)} \\ \vdots & \vdots \\ 1 & x_n^{(1)} \end{array}\right] & \boxed{n \times 1} \\ a_0 \qquad b_1 & \end{array}$$

$a_0, b_1, \ldots$

$\boxed{X = -2 : 0.01 : 2;}$ ✓

$\boxed{X = -2 + 4 \times \text{rand};}$ ✓

$$\rightarrow \; y_t = a_0 + b_1 x_t + \varepsilon_t$$

$\hat{y}_t = a_0 + b_1 x_t$ ← model data

y_t ← true observation

RMS

$$\sum_{t=1}^{T} \left(y_t - \hat{y}_t \right)^2$$

loss function

$$= \frac{1}{T} \boxed{\sum_{t=1}^{T} \left(y_t - a_0 - b_1 x_t \right)^2}$$

Goal: To miminize the loss function distane for all t's

$$\mathcal{L}(a_0, b_1) = \frac{1}{n} \sum_{j=1}^{n} \left(y_j - a_0 - b_1 x_j \right)^2$$

$$\nabla \mathcal{L} = \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial a_0} \\ \frac{\partial \mathcal{L}}{\partial b_1} \end{pmatrix} = \begin{pmatrix} -\frac{1}{n} \times 2 \sum_{j=1}^{n} \left(y_j - a_0 - b_1 x_j \right) \\ \frac{-2}{n} \sum^{n} \left(y_j - a_0 - b_1 x_j \right) x_j \end{pmatrix}$$

$j=1$

$\boxed{\text{Contour}}$ work up to 2 dimensional

+ 1

$d > 2$ $d=5$

$\boxed{\text{loss function}}$

38

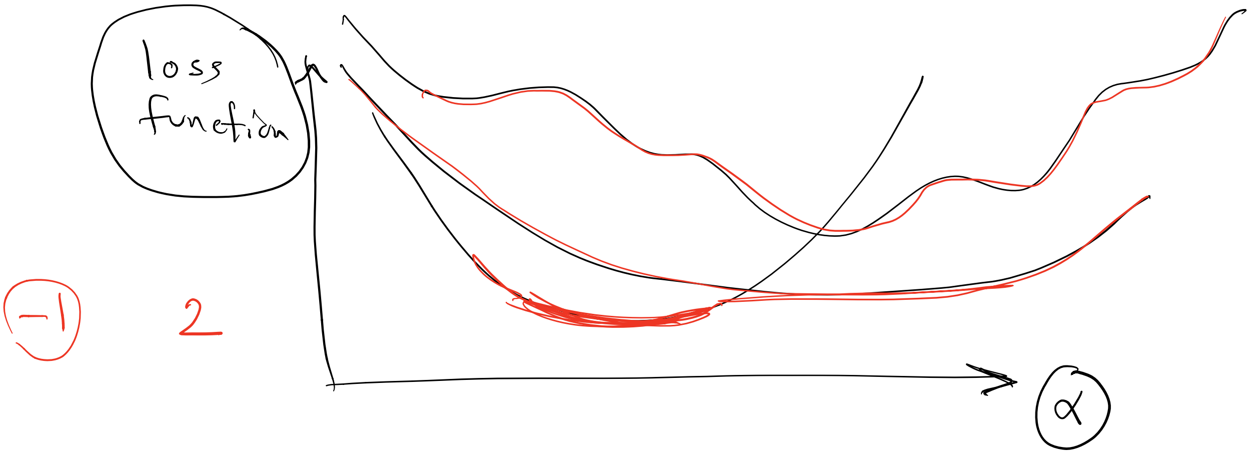
$$X_0 = [a_0; b_1; b_2; b_3; b_4]$$

$$X^* = [a_0^*; b_1^*; b_2^*; b_3^*; b_4^*]$$

$$\boxed{X_{temp} = \alpha X_0 + (1-\alpha) X^*}$$

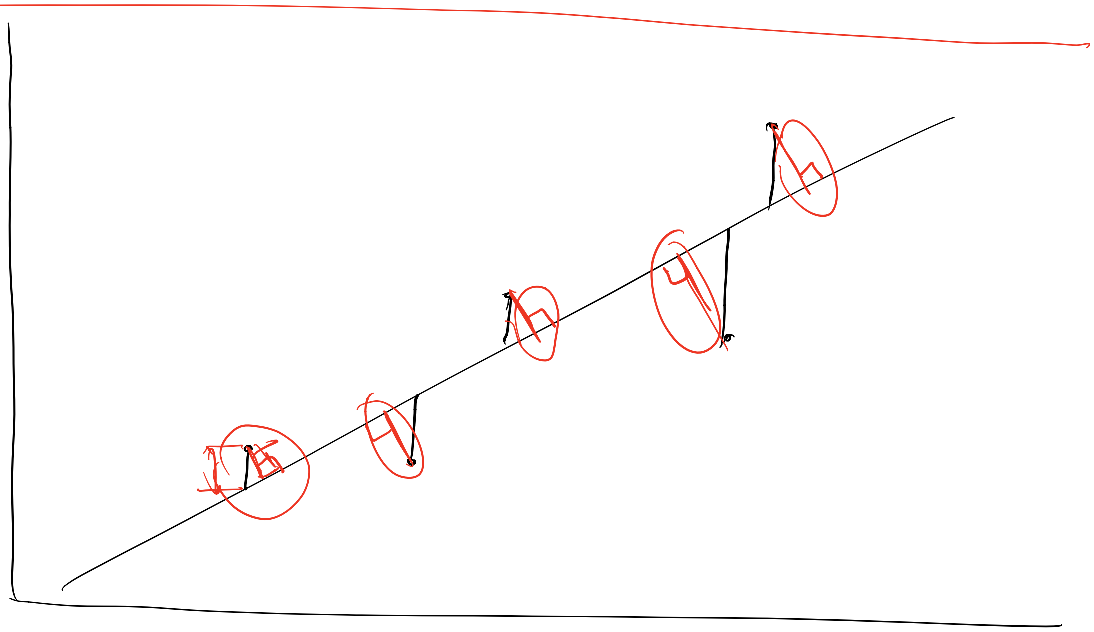
$\alpha = 1$ $\boxed{X_{temp} = X_0}$

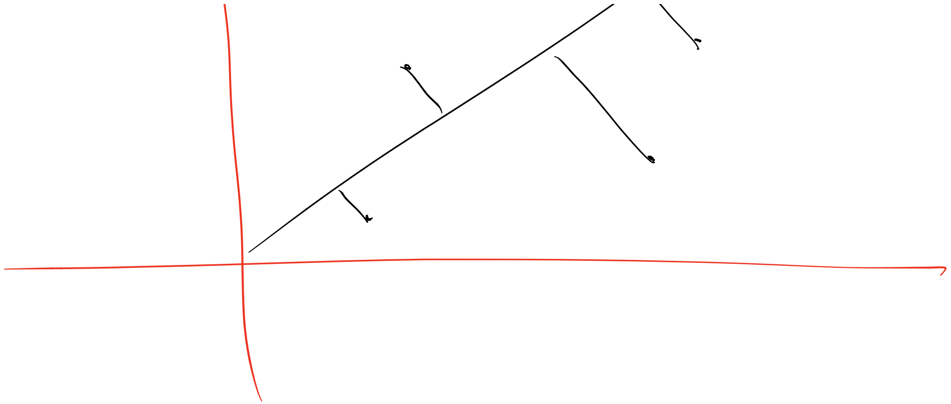
$\alpha = 0$ $\boxed{X_{temp} = X^*}$

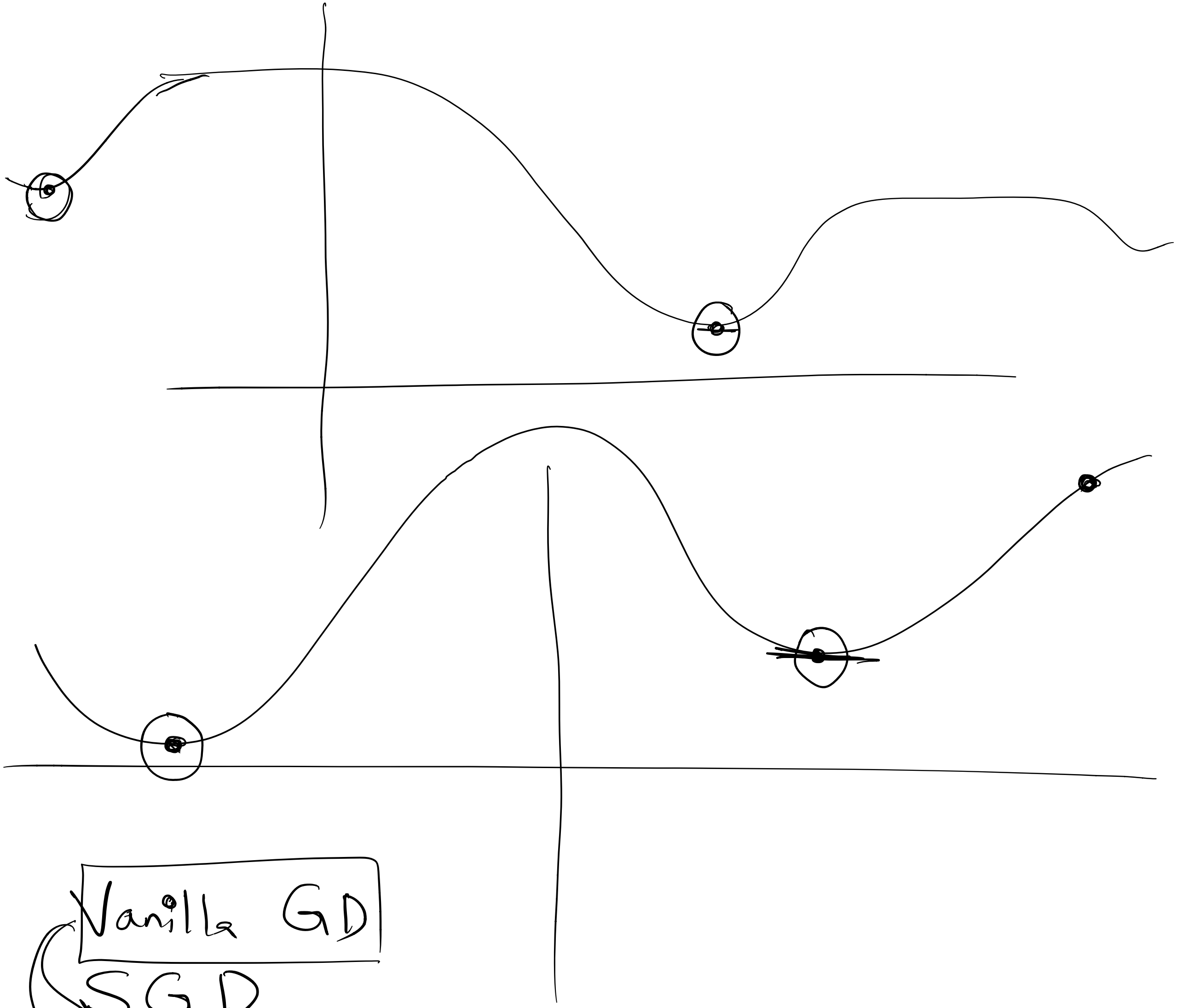


$X_0^{(1)}$ X^*

$X_0^{(2)}$







Vanilla GD

SGD

BGD

- Learning Algorithm
  - ✓ Statistical ML / Bayesian updates
  - ✓ ML
  - ✓ DL

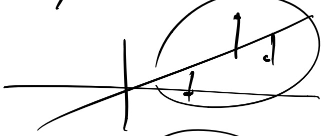
Loss functions & optimization

Loss functions

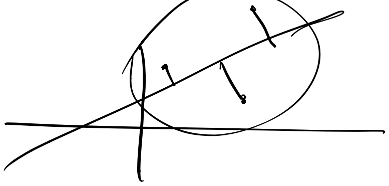
- Choice of starting point
- Choice of optimization

  Simplex vs. Gradient Descent

- Choice of objective function/loss function
  - distance



  - y difference



- Stress-testing ✓