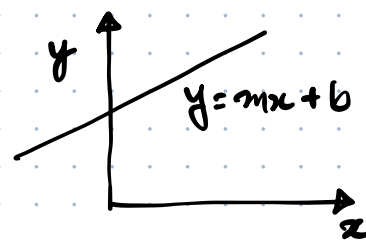


Oct 1, 2026

last week: neural networks

- classification
- regression

2D line fitting.



$$y = [x \ 1] \begin{bmatrix} m \\ b \end{bmatrix}$$
$$= x^T \theta$$

↑
model parameters

$$y = f(\vec{x}; \theta)$$

← non-linearity.

Estimate θ .

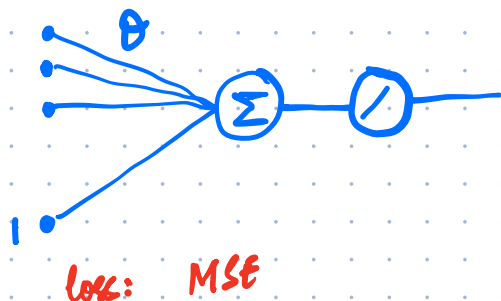
(i) Define loss L_θ

(ii) $\frac{\partial L_\theta}{\partial \theta}$

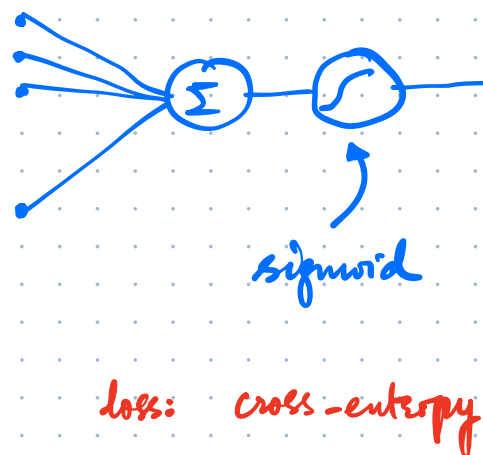
(iii) $\theta_{i+1} \leftarrow \theta_i - \lambda \frac{\partial L_\theta}{\partial \theta}$

← learning rate

Regression:



Classification:



$$\underbrace{f(\underbrace{g(x)}_{\Sigma})}_{\text{SOFTMAX}}$$

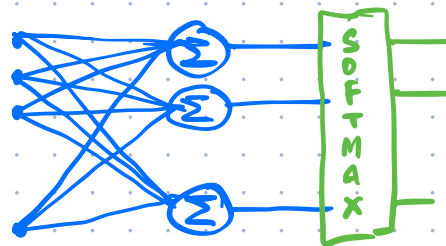
$$f(x)$$

$$\frac{\partial f}{\partial x}$$

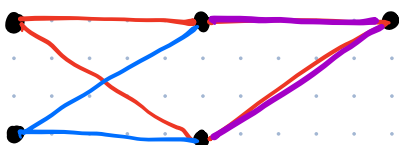
$$\frac{\partial f}{\partial x} = ?$$

$$= \frac{\partial f}{\partial g} \frac{\partial g}{\partial x}$$

CHAIN-RULE.

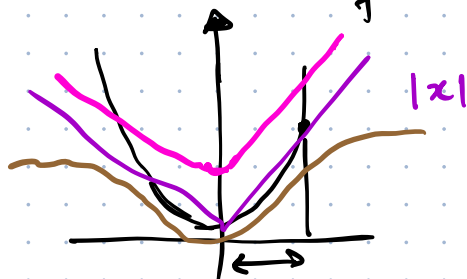


loss: cross-entropy

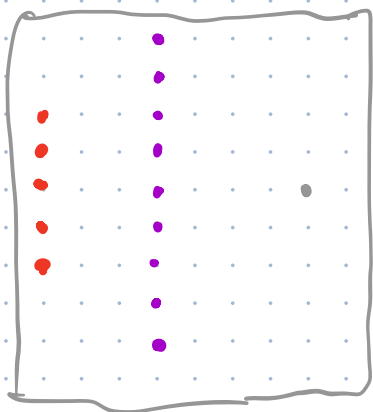


Q. Give me one drawback of using MSE loss?

A. Outliers.

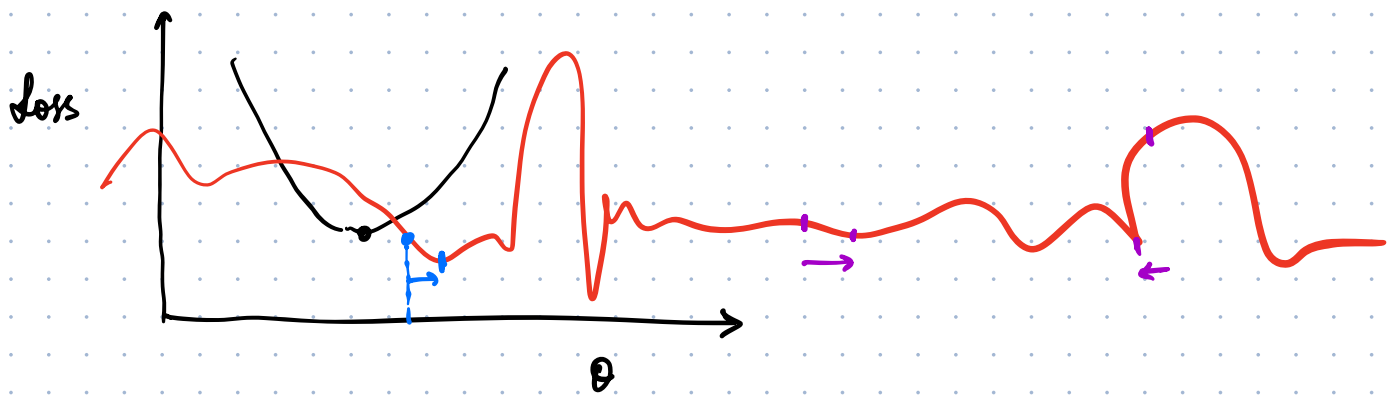


$$y = f_o(W_{oh} f_i(W_{ih} x))$$



$$W_{ih} \in \mathbb{R}^{9 \times 5}$$

$$W_{oh} \in \mathbb{R}^{1 \times 9}$$



$$y = mx + b \quad \leftarrow \text{Forward fn.}$$

$\uparrow$ $\uparrow$
 output input

①

$$\frac{\partial z_{l+1}}{\partial z_l} = \frac{\partial y}{\partial x} = m$$

②

$$\frac{\partial z_{l+1}}{\partial \theta} = \begin{cases} \frac{\partial y}{\partial m} = x \\ \frac{\partial y}{\partial b} = 1 \end{cases}$$