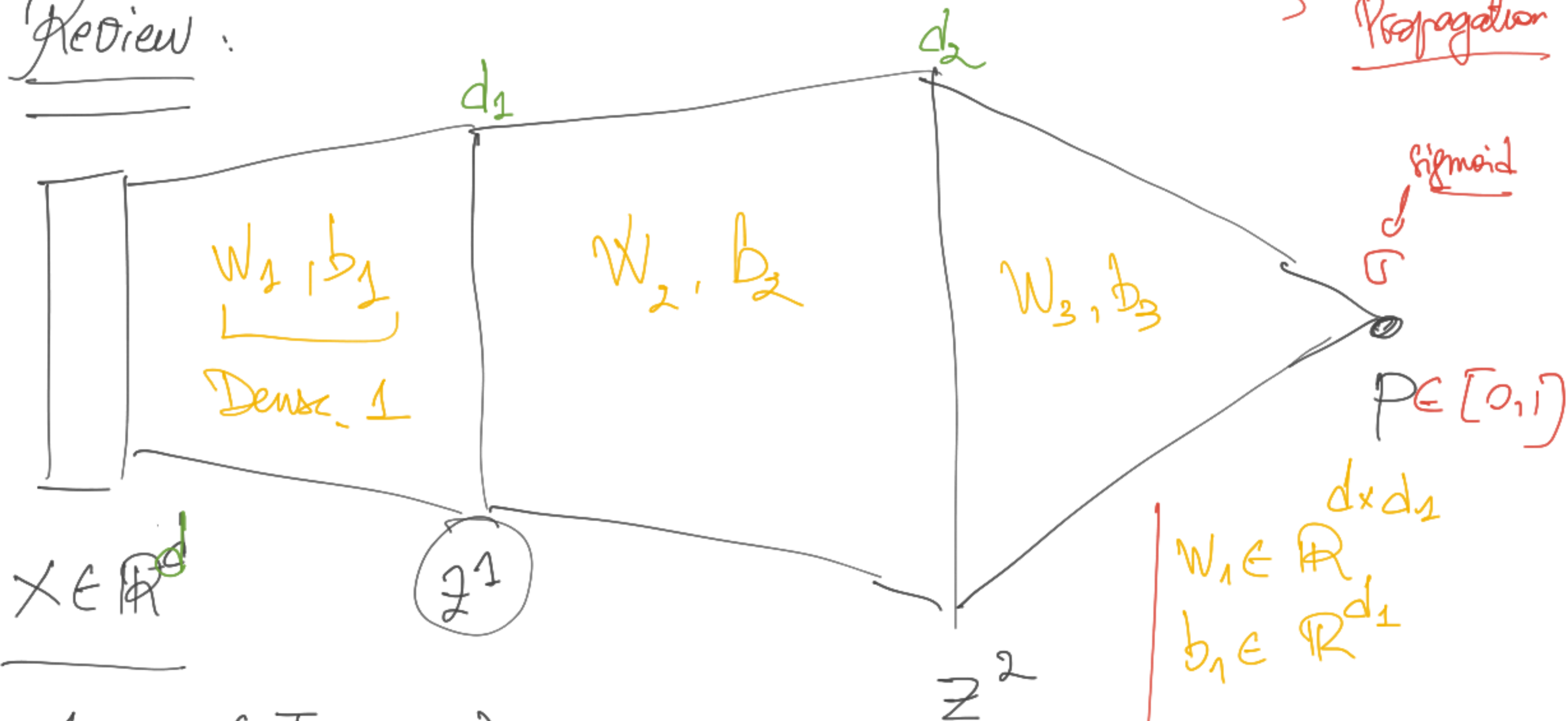


Review:




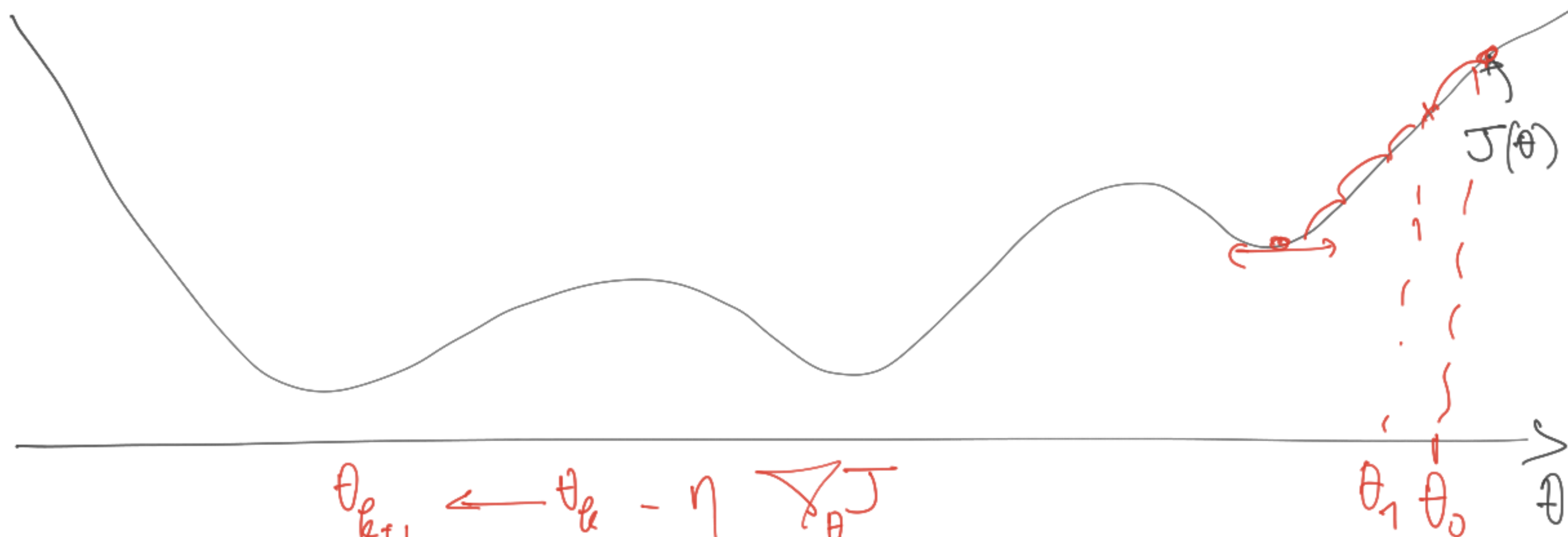
$$z^1 = \sigma(W_1^T x + b_1)$$

$$z^2 = \sigma(W_2^T z^1 + b_2)$$



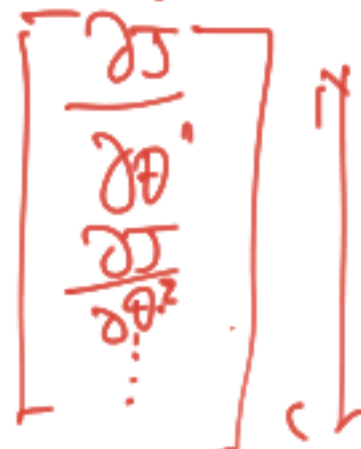

$$p = \sigma(W_3^T z^2 + b_3)$$

$$W_1 \in \mathbb{R}^{d \times d_1}$$
$$b_1 \in \mathbb{R}^{d_1}$$

Parameters : $\theta = \{w_i, b_i; 1 \leq i \leq 3\} =$  all parameters



$$\theta_{k+1} \leftarrow \theta_k - \eta \nabla_{\theta} J$$

$\sim 10^3$

parameters

Loss function: $J(\theta) = - \frac{\log(L(\theta))}{N}$

Es Binary Classification pb:

$$J(\theta) = -\frac{1}{N} \sum_{i=1}^N \{ y_i \log p_i + (1-y_i) \log (1-p_i) \}$$

Dataset : $\{x_i, y_i\}, 1 \leq i \leq N$

$$x_i \xrightarrow{f_{\theta}} p_i$$

$$p_i = \textcircled{f_{\theta}}(x_i)$$

NN

• Multiclass classification pb : (with K categories)

Dataset : $\{X_i, y_i\}_{i=1}^N$; $X_i \in \mathbb{R}^d$; $y_i \in \{1, \dots, K\}$
 $\hookrightarrow \hat{y}_i = [0 \dots 1 0 \dots]_K$

$$X_i \xrightarrow{\phi} \underline{P_i} = \begin{bmatrix} P_i^1 \\ \vdots \\ P_i^K \end{bmatrix}$$

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \hat{y}_i^k \log(P_i^k)$$

Dataset

doc₁: "Neural Networks are fun" \Rightarrow [—————]_v

doc₂: Attention models are awesome \Rightarrow [—————]_v

⋮

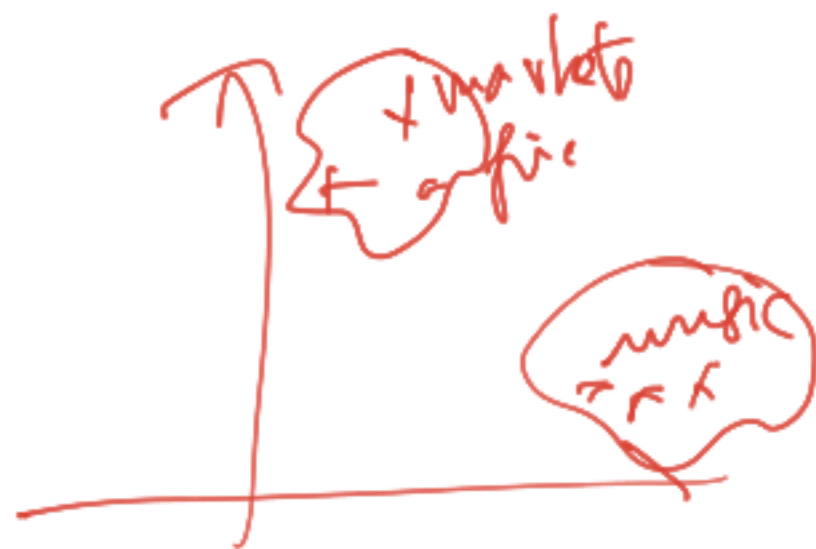
doc_N: This course is doing \Rightarrow [—————]_v

data processed is of shape (N, V)

Recap

GloVe approach

$D \sim 50 - 300$



$$\underline{W} = \begin{bmatrix} \text{---} w_1 \text{---} \\ \vdots \\ \text{---} w_V \text{---} \end{bmatrix} \in \mathbb{R}^{U \times D}$$

} D-dim representation of each word (as a center word) (~ users)

$$\underline{W}^2 = \begin{bmatrix} \text{---} w_1^2 \text{---} \\ \vdots \\ \text{---} w_V^2 \text{---} \end{bmatrix} \in \mathbb{R}^{U \times D}$$

} D-dim representation of each word (as a context word) (~ movies)

do

several Networks are fun.



\mathcal{E}_{32}



\mathcal{E}_{47}



\mathcal{E}_{100}



\mathcal{E}_2



Embedding Matrix (GloVe)

$$\mathcal{E} = \begin{bmatrix} \text{---} \mathcal{E}_{32} \text{---} \\ \text{---} \mathcal{E}_{47} \text{---} \\ \vdots \\ \text{---} \mathcal{E}_2 \text{---} \end{bmatrix} \begin{matrix} 1 \\ \vdots \\ r \end{matrix}$$

$\in \mathbb{R}^{U \times D}$

add

$$\rightarrow [\mathcal{E}_{32}, \mathcal{E}_{47}, \mathcal{E}_{100}, \mathcal{E}_2] \in \mathbb{R}^{T \times D}$$



doc₁

→ (T, D)

doc₂

→ (T, D)

⋮

doc_N

→ (T, D)

Dataset

processed

(N, T, D)