

Sentiment with Deep NN

I. Data processing

$$\text{Tweet: Train } X = \left[\begin{array}{l} [\dots \text{tweet}_1 \dots], \\ [\dots \text{tweet}_2 \dots], \\ [\dots \text{tweet}_3 \dots], \\ [\dots \text{tweet}_4 \dots], \\ [\dots \text{tweet}_5 \dots] \end{array} \right] \quad \text{Train } Y = \left[\begin{array}{l} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \right]$$

positive
negative

↳ Process tweet (tokenize) \Rightarrow count occurrence of word

\rightarrow Vocab dict. Ex. Vocab = {
 'PAD': 0,
 'UNK': 1,
 'Engage': 2,
 'top': 3,
 ... : 5555,
 ... : }

Convert a tweet to a tensor

change list of words to its index in vocab

Ex. tweet : 'I want to learn NLP' \Rightarrow [600, 7778, 5555, 448, 4475] (1)

Process batch

Ex. batch = 10 \Rightarrow $\left[\begin{array}{l} [\text{tensor of tweet}_1 \text{ (positive)} \dots], \\ [\dots \dots \dots \dots \dots \dots \dots \dots \dots \dots], \\ [\text{tensor of tweet}_2 \text{ (negative)} \dots], \\ [\dots \dots \dots \dots \dots \dots \dots \dots \dots \dots] \end{array} \right]$

\hookrightarrow Add padding so that all tensor has the same length.

\hookrightarrow Ex. get maximum length of tensor = 10 $\xrightarrow{\text{padding}}$ [0, 0, 0, 0, 0]

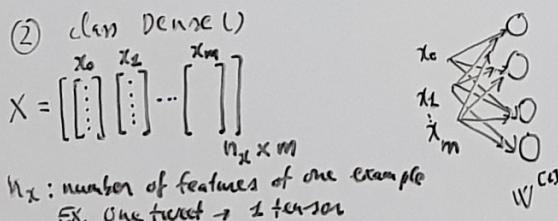
\hookrightarrow (1) \Rightarrow [600, 7778, 5555, 448, 4475, 0, 0, 0, 0, 0]

II. Implement

① class Relu()

$$\begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases} \quad \text{np.maximum}(z, 0) \text{ where } z \text{ is np array}$$

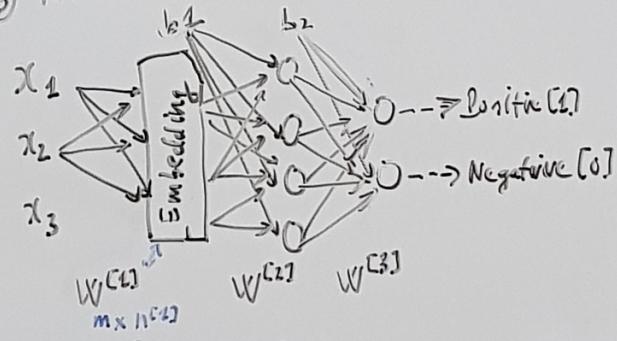
② class Dense()



$$\text{Forward: } X \cdot W \quad n_x \times m \quad m \times n^{(1)}$$

where $n^{(1)}$ number of units in Layer 1

③ NN for sentiment analysis



m : number of words \geq vocab-size

$n^{(1)}$: length of embedding for each word

Model = Sequential([
 embedding-layer, $\# W^{(1)}$
 mean-layer, $\#$ average of all word for each features. $[1]_{n \times 1}$
 dense-layer, $\# W^{(2)}$ shape $(1 \times n^{(2)} \text{ units})$
 (log-softmax-layer) $\# W^{(3)}$ shape (2×1)])