University of BRISTOL

Neural networks: Chapter 7, Speech and Language Processing, 3rd edition draft, Jurafsky & Martin (2021).

8.1 Deep Text Classifiers

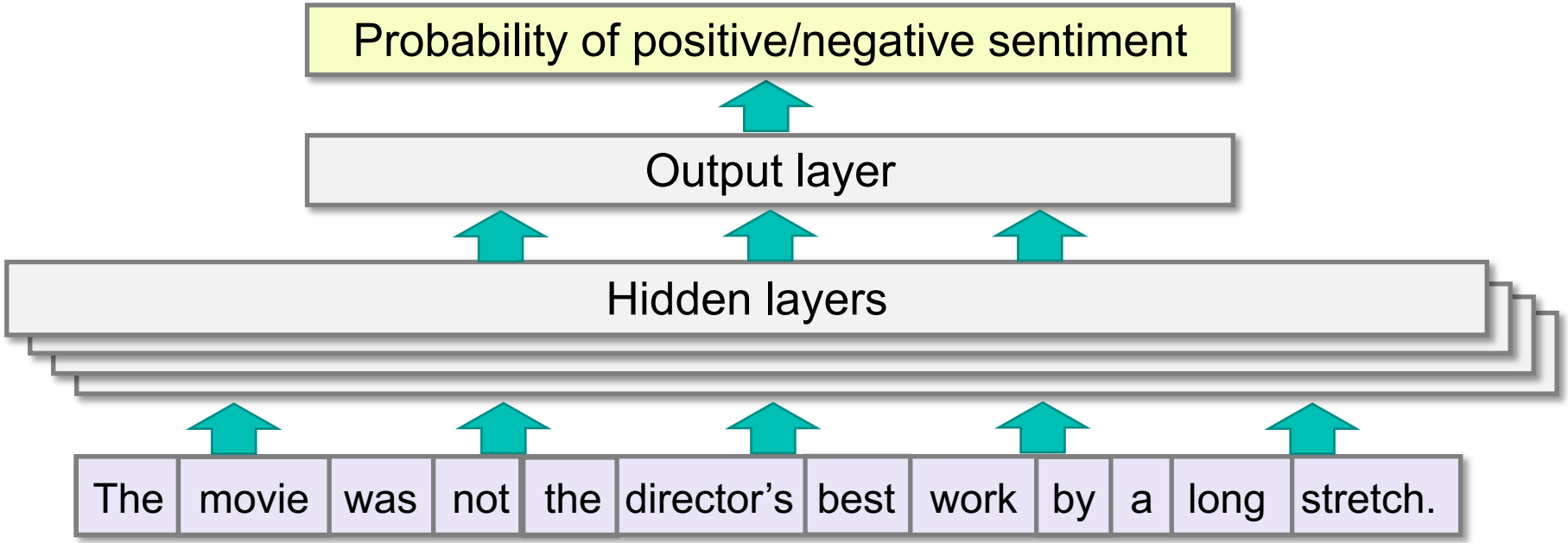
Edwin Simpson

Department of Computer Science,

University of Bristol, UK.

bristol.ac.uk

Neural Network (NN) Text Classifier

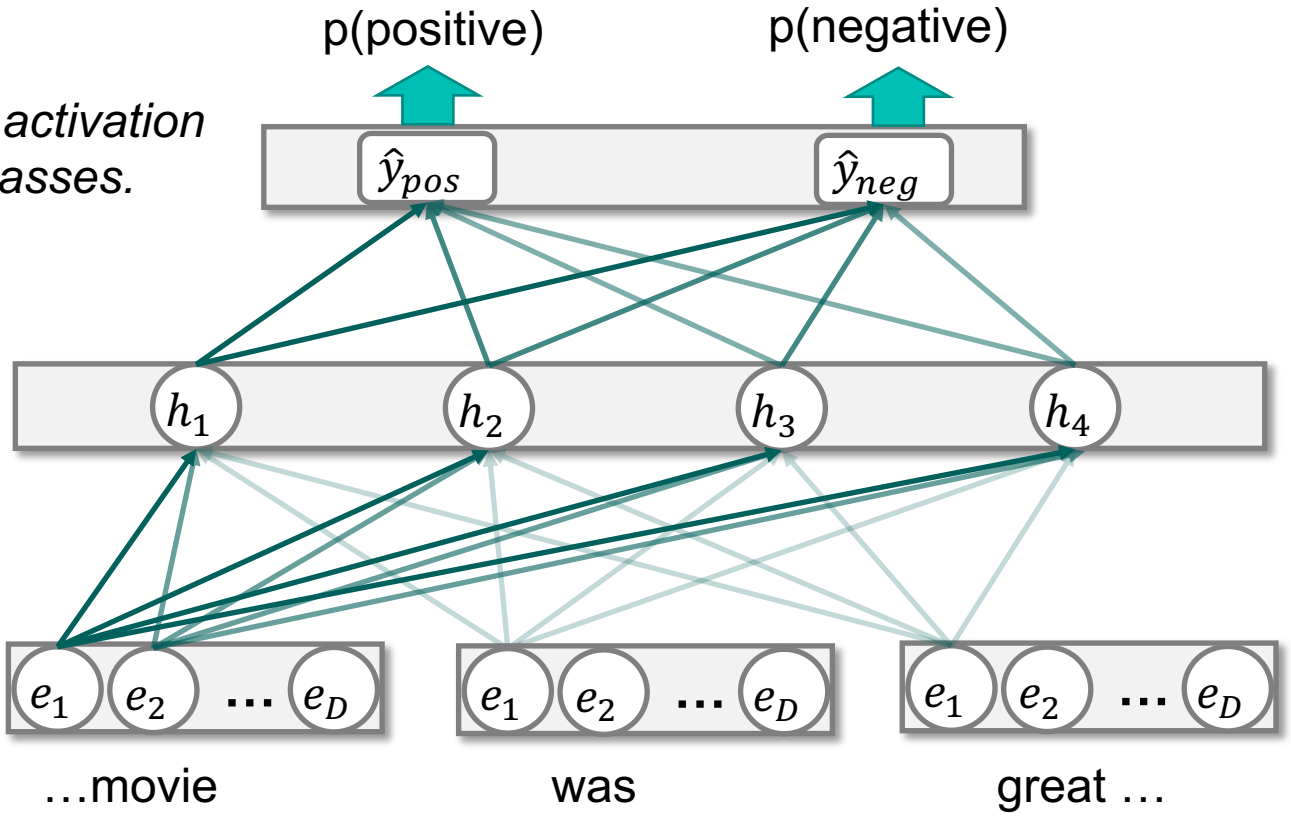


Neural Text Classifier

Output layer: softmax activation function for multiple classes.

Fully-connected Hidden layer: ReLU activation function.

Input layer: embeddings or one-hot encoding



Fine-tuning Embeddings

- Word embeddings can be stored in a |V| x D matrix, *E*:
- Look up a word's embedding by multiplying its one-hot encoding with the matrix.
- *E,* can be treated as another weight vector in an embedding layer.
- Then we can train the embedding matrix through backpropagation.
  - Embeddings are **pretrained** using a model like Skipgram;
  - Then **fine-tuned** on the training set to better represent the features that are important for this task.
  - This is **end-to-end** deep learning: the whole pipeline sits inside one model.

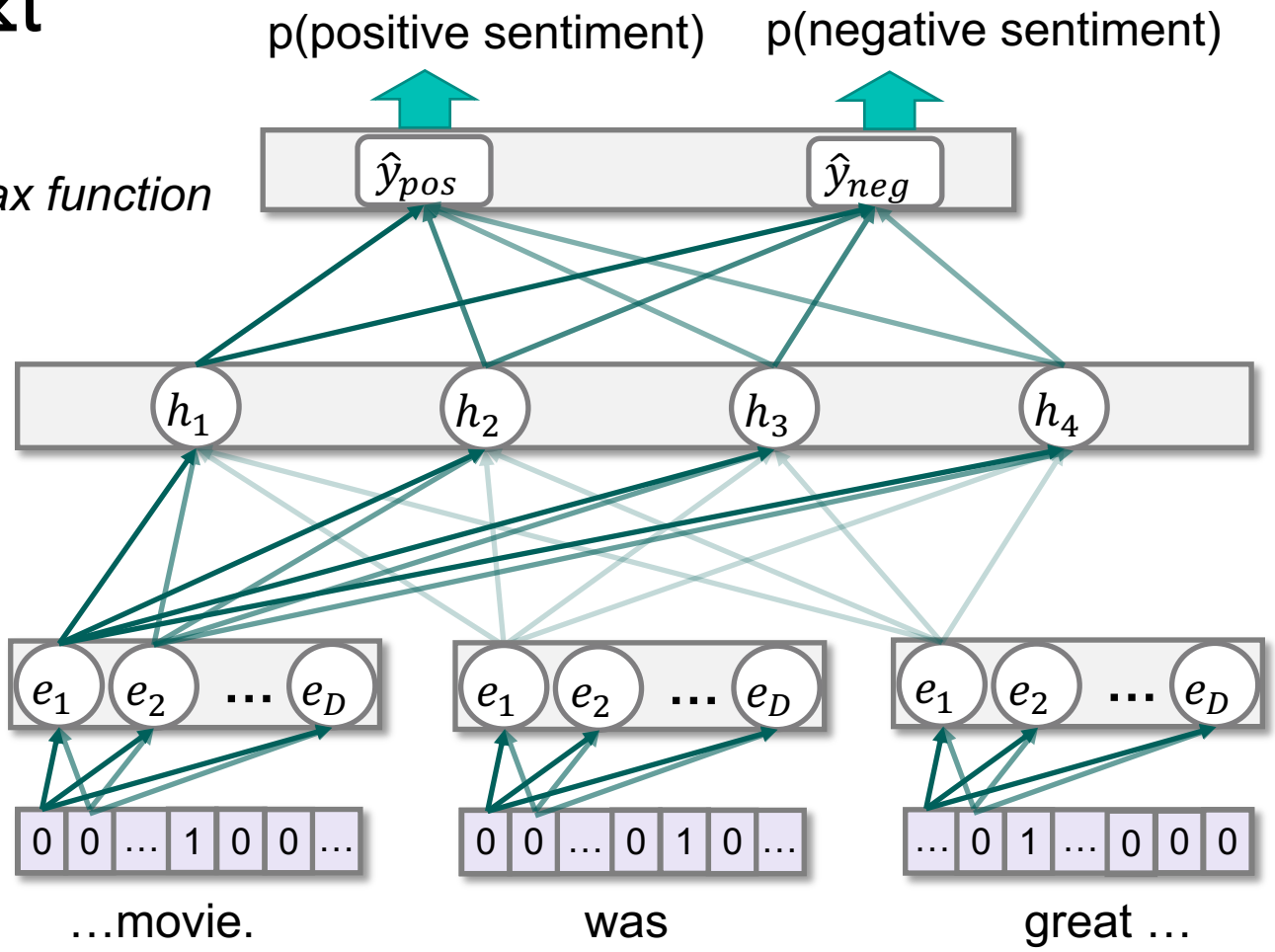
Neural Text Classifier

Output layer: softmax function

Feed-forward Hidden layer

Embedding layer

Input layer: one-hot encoding



Padding

- The weight matrix $\boldsymbol{W}^{(l)}$ for each layer has a fixed size.
- But the input texts can be of varying lengths.
- Define a maximum document length N and truncate long documents.
- **Pad** short documents by appending special '[PAD]' tokens to the sequence until it has length N**.**

Summary

- Neural networks can process sequences of words for tasks like text classification.
- Inputs can be either one-hot encodings or fixed word embeddings.
- If we include an embedding layer, the embeddings can be fine-tuned to the training set.
- Padding is needed to convert documents of varying lengths to a fixed-size input.