ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI FORLÌ

91258 / B0385
Natural Language Processing

Lesson 20. Into Transformers[1]

Alberto Barrón-Cedeño
a.barron@unibo.it

10/12/2025

[1] Partially based on `medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04`; not offered in 2025 because students preferred a Q&A session

Table of Contents

Generating text with LSTM[2]

[2] In the 2025 edition of the lesson, we closed with this topic and a Q&A

Influencing the level of *creativity*

The output layer is a softmax; i.e. it produces a probability distribution

We do not select the token with the highest score, but randomly predict from the distribution

Dividing the log by the temperature modifies such distribution:

- < 1 sharpens the distribution $\rightarrow$ strict attempt to recreate the original text
- > 1 flattens the distribution $\rightarrow$ attempts to produce a more diverse output

Let us look at function sample

Influencing the level of *creativity*

$$[-1.20, -0.69, -1.61] = log([0.3, 0.5, 0.2]) \quad (1)$$

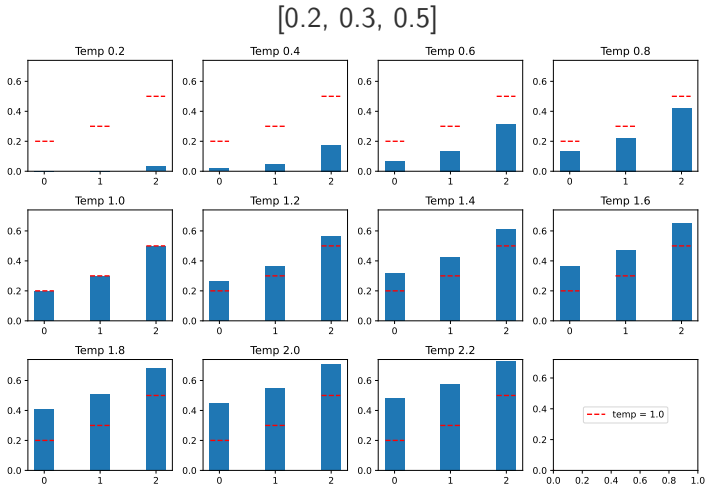
temperature = 1.2

$$\begin{aligned} [-1.00, -0.58, -1.34] &= \frac{[-1.20, -0.69, -1.61]}{1.2} \\ [0.37, 0.56, 0.26] &= exp([-1.00, -0.58, -1.34]) \end{aligned}$$

temperature = 0.8

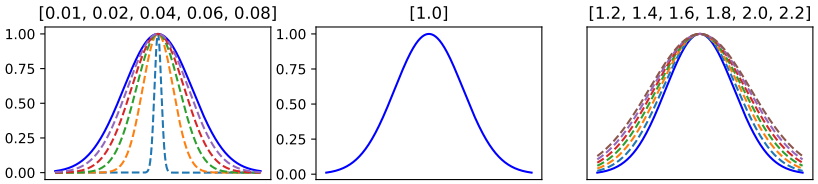
$$\begin{aligned} [-1.51, -0.87, -2.01] &= log([0.3, 0.5, 0.2])/0.8 \\ [0.22, 0.42, 0.13] &= exp([-1.51, -0.87, -2.01]) \end{aligned}$$

Influencing the level of *creativity*



Influencing the level of *creativity*

Looking at a Gaussian distribution



Q&A

Recap

Recap: The path

1. Baby steps into computing
2. What is NLP? From rule-based to statistical
3. Pre-processing text: tokens, stemming, stopwording...
4. From words to vectors: the vector space model
5. A few supervised models
6. Training and evaluating in machine learning
7. From words to meaning: topic modeling
8. Using one *neuron*: perceptron
9. Fully-connected neural networks
10. From words to semantics: word embeddings
11. Taking snapshots of text: CNNs
12. Texts as sequences: (Bi)RNNs
13. Using a better memory: LSTM
14. LSTM to produce text
15. Intro to transformers

Recap: The future path

- We covered Parts 1 and 2 of Lane et al. (2019) (up to Section 9)
- That's 9 out of 13 chapters of Natural Language Processing in Action

  Now go and celebrate the end of the course
  ...and worry about your project from Jan 2nd!

- I'm available during January for 1-to-1 discussion on your project upon request!

References I

Lane, H., C. Howard, and H. Hapkem
2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning Publication Co.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin
2017. Attention is all you need.