Neural Probabilistic Language Model (NPLM)

Index

- NPLM structure
- NPLM and its meaning information

Fundamental Structure

- Based on the language model. (+ overcoming its limitations)
- ‘n-gram’ language model
- Using softmax as training method

Output

Softmax function

- $P(w_t|w_{t-1}, \dots, w_{t-1}, w_{t-n+1}) = \frac{\exp(y_{w_t})}{\sum_i \exp(y_i)}$

  y_{w_t} : score vector of w_t, V(the # of words) dimensions
- Raise the nominator, lower the denominator
  → increase the conditional probability of P
- Back propagation with target index.

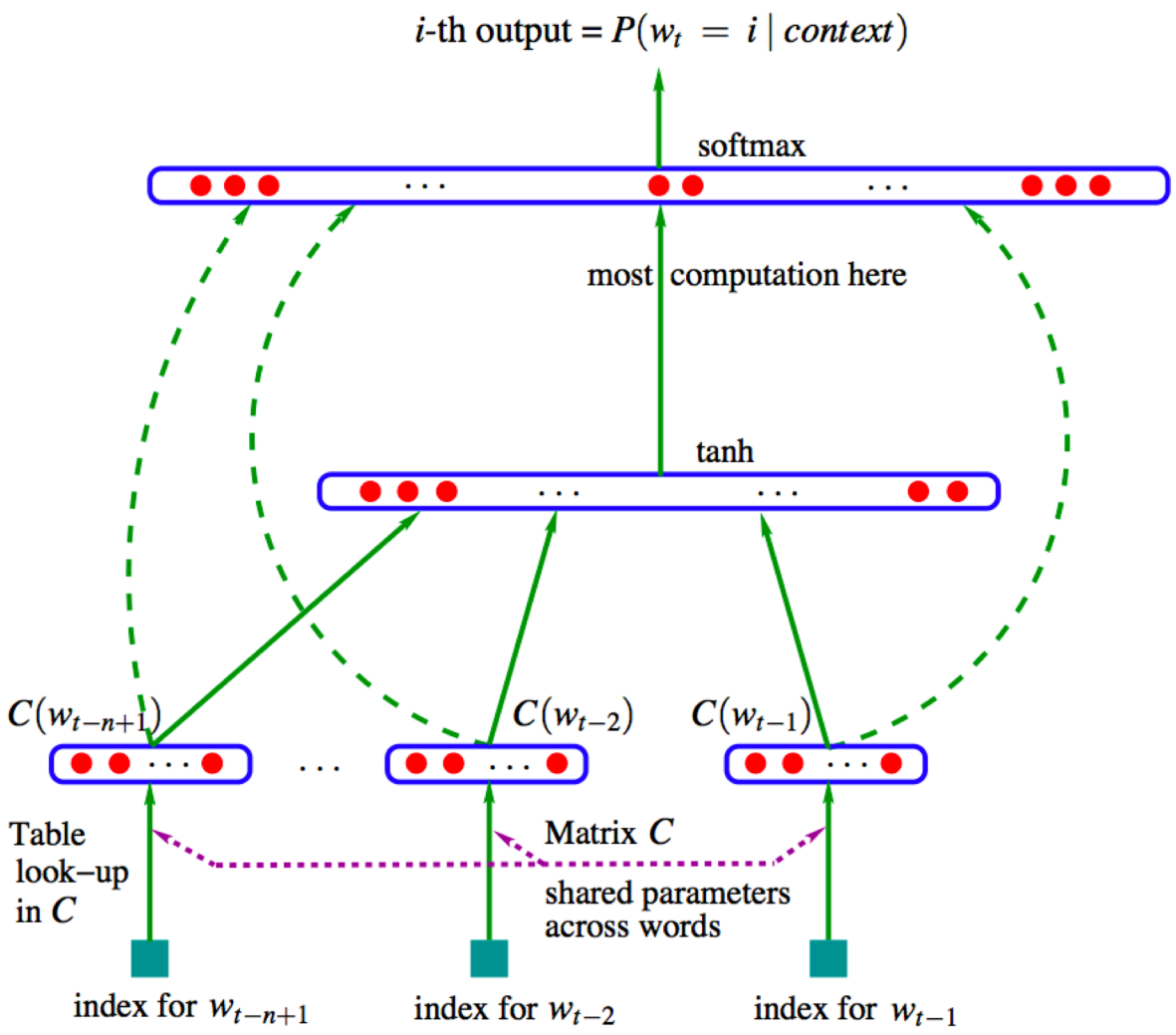
Input

- $x_t = \mathrm{C}(w_t)$

$\mathrm{C}(w_t)$: inner product of matrix C and vector w_t

w_t : One-hot-vector, C : $|V| \times m$ matrix

- $x = [x_{t-1}, x_{t-2}, \dots, x_{t-n+1}],\ x \in \mathbb{R}^{(n-1)m}$
- $y_{w_t} = b + Wx + U \tanh(d + Hx),\ y \in \mathbb{R}^{|V|}$



NPLM and its information

- Row vectors of Matrix C is updated by receiving the gradient which minimize the train loss occurred during training.

  → the word vectors moved to same direction in vector space.

  → Matrix C contains the context info of each word.

- Too many parameters. ($H,\ x,\ d,\ W,\ U,\ b,\ y,\ C$)

  →Word2Vec