Neural network visualization

9

Neural network decision boundary visualization

Neural network visualization

`Neural network decision boundary visualization`

Neural language modeling

- In n-gram modeling, we take the conditional probability of the next word:

$$P(\mathbf{x}^{(t)} \mid \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(t-1)}) \quad (1)$$

- And because of data sparsity must apply the Markov assumption, yielding:

$$P(\mathbf{x}^{(t)} \mid \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(t-1)}) = P(\mathbf{x}^{(t)} \mid \mathbf{x}^{(t-n)}, \ldots, \mathbf{x}^{(t-1)}) \quad (2)$$

- Neural language modeling is much more flexible and more powerful:
    - Unlike n-grams, NNs can directly model infinite context windows (eq 1) using one of two methods:
        - **Convolutional filtering**
        - **Recurrent network connections**
    - Like word2vec, unlike n-grams, learns a distributed representation your tokens
    - Unlike word2vec and n-grams which are linear models, NNs can model arbitrarily complex, highly non-linear relationships between tokens in a sentence!