1 Independence Assumptions in Log-linear Taggers

1.1 Question (time: 8:32, slide: 8)

Say we have $w_1 \dots w_3$ = the dog barks. We would like

- $p($ D N V | the dog barks$) = 0.5$
- $p($ D N N | the dog barks$) = 0.5$

What should be the value for the following probabilities

- $p($ D | the dog barks, * *$)$
- $p($ N | the dog barks, * D$)$
- $p($ V | the dog barks, D N$)$
- $p($ N | the dog barks, D N$)$

Write your answers to one decimal place separated by a space, e.g. 0.0 0.9 1.0 0.1

2 Features in Log-Linear Taggers

2.1 Question (time: 3:41, slide: 10)

Say we are given the following sentence with partial tag sequence

- the/DT mat/NN saw/VBD the/DT dog/NN in the park

Which of the following tuples is the correct history at this point?

(a) (DT, NN, the man saw the dog, 6)

(b) (NN, DT, the man saw the dog, 6)

(c) (DT, NN, the man saw the dog in the park, 6)

(d) (NN, IN, the man saw the dog in the park, 6)

2.2 Question (time: 13:21, slide: 14)

Say we are given a history

- $h =$ (VBD, DT, the man saw the dog in the park, 5)

Which of the following features have $f(h, \text{NN}) = 1$?

(a) $f(h,t) = \begin{cases} 1 & \text{if } w_i = \text{ dog and } t = \text{ NN} \\ 0 & \text{otherwise} \end{cases}$

(b) $f(h,t) = \begin{cases} 1 & \text{if } w_i = \text{ dog and } t = \text{ DT} \\ 0 & \text{otherwise} \end{cases}$

(c) $f(h,t) = \begin{cases} 1 & \text{if } w_{i+1} = \text{ dog and } t = \text{ NN} \\ 0 & \text{otherwise} \end{cases}$

(d) $f(h,t) = \begin{cases} 1 & \text{if } y_{-1} = \text{ DT and } t = \text{ NN} \\ 0 & \text{otherwise} \end{cases}$

3 Parameters in Log-Linear Taggers

3.1 Question (time: 3:59, slide: 16)

Consider the training example

- the dog saw the man, DT NN VBD DT NN

Say we convert this example to (x_i, y_i) pairs for training a log-linear model. Which of the following items are in the training set?

(a) $x =$ (DT NN, the dog saw the man, 3), y = NN

(b) $x =$ (VBD DT, the dog saw the man, 3), y = VBD

(c) $x =$ (DT NN, the dog saw the man, 3), y = VBD

(d) $x =$ (DT NN, the dog saw the man, 4), y = NN

4 The Viterbi Algorithm for Log-linear Taggers

4.1 Question (time: 9:37, slide: 20)

Consider the sentence $w_1 \ldots w_n =$ the dog saw the cat.

Say we have the π values

π(3, NN, VBD) = 0.02 π(3, NN, NN) = 0.01 π(3, NN, DT) = 0.03

π(3, u, v) = 0.0 for all other values of u and v And say our log-linear model gives

- q(DT | NN, VBD, the dog saw the cat, 4) = 0.4

- $q(\text{DT} \mid \text{NN}, \text{NN}, \text{the dog saw the cat}, 4) = 0.5$
- $q(\text{DT} \mid \text{DT}, \text{NN}, \text{the dog saw the cat}, 4) = 0.3$

What is the value for $\pi(4, \text{VBD}, \text{DT})$?

A Answers

- 1.0 1.0 0.5 0.5

  The two taggings, D N V and D N N, differ only in the last tag and are the only possible taggings for this sentence (probability sums to 1). Therefore we set the last decision probabilities to 0.5 each. The correct answer is 1.0 1.0 0.5 0.5.

- (c)

  The incorrect answers either do not have the full sentence in the history or have the part-of-speech tags in the wrong order.

- (a) (d)

  The incorrect answers either have $t \neq$ NN or have w_{i+1} equal to the wrong word.

- (c)

  The incorrect responses either have y not equal to the tag at position i or have t_{i-1} or t_{i-2} not equal to the same tags as in the correct tagging.

- 0.008

  The answer is 0.008. The value for $\pi(4, \text{VBD}, \text{DT}) = \max_x q(\text{DT} \,|x, \text{VBD}, \text{the dog saw the cat}, 4) \times \pi(3, x, \text{VBD}) = q(\text{DT} \mid \text{NN}, \text{VBD}, \text{the dog saw the cat}, 4) \times \pi(3, \text{NN}, \text{VBD}) = 0.008$.