1 Three Components of GLMs

1.1 Question (time: 8:46, slide: 17)

Consider a set of tags $\mathcal{T} = \{$ DT, V, NN $\}$.

Say we are given the sentence $x =$ `the dog walked to the store` and asked to construct a global linear model for tagging.

What is an upper bound on the size of GEN(x)?

1.2 Question (time: 8:46, slide: 17)

Consider a set of tags $\mathcal{T} = \{$ DT, V, NN $\}$.

Say we are given the sentence $x =$ `the dog walked to the store` and asked to construct a global linear model for tagging.

If we define GEN(x) to generate the top N tag sequences where $N = 100$, what is the size of GEN(x)?

1.3 Question (time: 14:39, slide: 0)

Consider $x =$ `the dog laughs` and GEN(x) made up of the tag sequences

DT NN V   DT NN DT   DT DT DT

Now say we are given f consisting of the following three feature functions

- $f_1(x, y) = \begin{cases} 1 & \text{if } y \text{ starts with DT and ends with V} \\ 0 & \text{otherwise} \end{cases}$
- $f_2(x, y) = \begin{cases} 1 & \text{if } y \text{ contains two DTs} \\ 0 & \text{otherwise} \end{cases}$
- $f_3(x, y) = \begin{cases} 1 & \text{if all tags the same in } y \\ 0 & \text{otherwise} \end{cases}$

What is the feature vector of the first tagging in GEN(x)? (Write each value in the vector separated by a space, e.g. 1 1 0).

1.4 Question (time: 14:39, slide: 20)

Consider $x =$ `the dog laughs` and GEN(x) made up of the tag sequences

DT NN V   DT NN DT   DT DT DT

Now say we are given f consisting of the following three feature functions

- $f_1(x,y) = \begin{cases} 1 & \text{if } y \text{ starts with DT and ends with V} \\ 0 & \text{otherwise} \end{cases}$
- $f_2(x,y) = \begin{cases} 1 & \text{if } y \text{ contains two DTs} \\ 0 & \text{otherwise} \end{cases}$
- $f_3(x,y) = \begin{cases} 1 & \text{if all tags the same in } y \\ 0 & \text{otherwise} \end{cases}$

If we are given the weight vector $\mathrm{v} = \langle 10, 2, 9 \rangle$, what is $\max_{y \in \mathrm{GEN}(x)} f(x,y) \cdot v$?

2 Parameter Estimation with the Perceptron Algorithm

2.1 Question (time: 6:11, slide: 30)

Say we are running the perceptron algorithm. We have reached input x_i and the set $\{f(x_i, y) : y \in \mathrm{GEN}(x_i)\}$ is made up of the vectors

- $\langle 0, 1, 0, 1 \rangle$
- $\langle 0, 1, 1, 1 \rangle$
- $\langle 1, 1, 0, 1 \rangle$

Also we know that $f(x_i, y_i) = \langle 1, 1, 0, 1 \rangle$ and that our current parameters are $\mathrm{v} = \langle -2, 5, 2, 0 \rangle$.

What will be the value of v at the end of this iteration? (Write each value in the vector separated by a space, e.g. 0 1 1 0).

2.2 Question (time: 6:11, slide: 30)

Say we are running the perceptron algorithm. We have reached input x_i and the set $\{f(x_i, y) : y \in \mathrm{GEN}(x_i)\}$ is made up of the vectors

- $\langle 0, 1, 0, 1 \rangle$
- $\langle 0, 1, 1, 1 \rangle$
- $\langle 1, 1, 0, 1 \rangle$

Also we know that $f(x_i, y_i) = \langle 1, 1, 0, 1 \rangle$ and that our current $\mathrm{v} = \langle 2, 5, 1, 0 \rangle$.

What will be the value of v at the end of this iteration? (Write each value in the vector separated by a space, e.g. 0 1 1 0).

A Answers

- 729

  The answer is 729. At its largest, GEN(x) is a set containing all possible tag sequences. There are 3 tags and 6 words, which gives GEN(x) = 729.

- 100

  The answer is 100. GEN(x) is a set containing only the top 100 possible tag sequences. Even though there are 729 different tag sequences, we only consider the top 100 within GEN(x).

- 1 0 0

  The answer is 1 0 0. The sentence is the/DT dog/NN laughs/V. Of the three features only the first is 1, and the other two are 0.

- 11

  The answer is 11. The last sentence $y =$ DT DT DT has feature vector $f(x, y) = \langle 0, 1, 1 \rangle$, and so $f(x, y) \cdot v = 11$.

- -1 5 1 0

  The answer is -1 5 1 0. First we compute the highest scoring vector z_i which is $f(x_i, z_i) = \langle 0, 1, 1, 1 \rangle$. Then we update the parameters $\mathrm{v} = \mathrm{v} + f(x_i, y_i) - f(x_i, z_i) = \langle -1, 5, 1, 0 \rangle$

- 2 5 1 0

  The answer is 2 5 1 0. First we compute the highest scoring vector z_i which is $f(x_i, z_i) = \langle 1, 1, 0, 1 \rangle$. Then we update the parameters $\mathrm{v} = \mathrm{v} + f(x_i, y_i) - f(x_i, z_i) = \langle 2, 5, 1, 0 \rangle$. Since the correct answer has the same feature vector as the selected answer, the parameters do not change.