Anish Sachdeva
DTU/2K16/MC/13

Class Fun
Date
Page No.

NLP Class Assignment
Laplacian Smoothing

Corpus:

i) <s> This is a nice day <e>
ii) <s> This day is really nice, really nice it is <e>

Unigrams	Count
This	2
is	3
a	1
nice	3
day	2
<s>	2
<e>	2
really	2
it	1

$|V| = 9$

Bigrams	Count
<s> This	2
This is	1
is a	1
a nice	1
nice day	1
day <e>	1

Class Fun

Date

Page No.

day is 1
is really 1
really nice 2
nice really 1
nice it 1
it is 1
is <e> 1
|V| = 13

Word with highest probability for start token <s>

$$P(\text{This} \mid \langle s \rangle) = \frac{C(\langle s \rangle \text{ This}) + 1}{C(\langle s \rangle) + V}$$

$$\frac{2+1}{13+2} = \frac{3}{15} = \frac{1}{5}$$

<s> This

Now, highest probability for $P(x \mid \text{This})$

$P(x \mid \text{This})$, x = is

$$P(\text{is} \mid \text{this}) = \frac{C(\text{this, is}) + 1}{C(\text{this}) + V} = \frac{1+1}{13+2} = \frac{2}{15}$$

<s> This is ....

$P(x \mid \text{is})$ → to maximize

3 options x = a.

CLASS FUN

Date | Page No.

$$P(a/is) = \frac{C(is,a)+1}{C(is)+V} = \frac{1+1}{13+3} = \frac{2}{16} = \frac{1}{8}$$

Sentence: <s> This is a

$P(x|a)$ $x = nice$

$$P(nice/a) = \frac{C(a,nice)+1}{C(a)+V} = \frac{1+1}{13+1} = \frac{2}{14} = \frac{1}{7}$$

Sentence = <s> This is a nice

$P(x|nice)$ $x = it$

$$P(it/nice) = \frac{C(nice,it)+1}{C(nice)+V} = \frac{1+1}{13+3} = \frac{2}{16} = \frac{1}{8}$$

Sentence: <s> This is a nice it

$P(x|it)$, $x = is$

$$P(is/it) = \frac{C(it,is)+1}{C(it)+V} = \frac{1+1}{13+1} = \frac{2}{14} = \frac{1}{7}$$

~~P(x|~~ Sentence: <s> This is a nice it is

$P(x|is)$: We have 3 options (is a) (is, really), (is, <e>) We randomly select (is, really)

Class Fun

Date

Page No.

$P(\text{really} \mid \text{is}) = \frac{C(\text{is, really}) + 1}{C(\text{is}) + V} = \frac{1+1}{13+3}$

$= \frac{2}{16} = \frac{1}{8}$

Sentence: <s> This is a nice it is really ...

Now, we need to find $P(x \mid \text{really})$

We take x = ~~really~~ nice

$P(\text{nice} \mid \text{really}) = \frac{C(\text{really, nice}) + 1}{C(\text{really}) + V}$

$= \frac{2+1}{13+2} = \frac{3}{15} = \frac{1}{5}$

Sentence: <s> This is a nice, it is really nice

We now need to minimize $P(x \mid \text{nice})$

We have 3 options (nice, day) (nice, really) (nice, it)

We randomly select (nice, day)

$P(\text{day} \mid \text{nice}) = \frac{C(\text{nice, day}) + 1}{C(\text{nice}) + V} = \frac{1+1}{13+3} = \frac{2}{16}$

$= \frac{1}{8}$

Class Fun

Date | Page No.

Sentence:

<S> This is a nice it is really nice day .....

We need to maximize P(x | day)

We have 2 options (day, is) (day <e>)
We randomly select (day, <e>)

$$P(\langle e \rangle \mid day) = \frac{C(day\ \langle e \rangle) + 1}{C(day) + V} = \frac{1+1}{13+2}$$

$$= \frac{2}{15}$$

Sentence:
<S> This is a nice it is really nice day <e>

Since, we have come to an end, the language generation will stop.

In the above program, the computer program is taking a random value when words with equal probability are available.