

Count!

N-gram language models

Language modeling

This is the ...

house

rat

did

malt

What's the probability of the next word?

$$p(\text{house} \mid \text{this is the}) = ?$$

Toy corpus

This is the house that Jack built.

This is the malt

That lay in the house that Jack built.

This is the rat,

That ate the malt

That lay in the house that Jack built.

This is the cat,

That killed the rat,

That ate the malt

That lay in the house that Jack built.

$$p(\text{house} \mid \text{this is the}) =$$

Toy corpus

This is the house that Jack built.

This is the malt

That lay in the house that Jack built.

This is the rat,

That ate the malt

That lay in the house that Jack built.

This is the cat,

That killed the rat,

That ate the malt

That lay in the house that Jack built.

$$p(\text{house} \mid \text{this is the}) =$$

Toy corpus

This is the house that Jack built.

This is the malt

That lay in the house that Jack built.

This is the rat,

That ate the malt

That lay in the house that Jack built.

This is the cat,

That killed the rat,

That ate the malt

That lay in the house that Jack built.

$$p(\text{house} \mid \text{this is the}) =$$

Toy corpus

This is the house that Jack built.

This is the malt

That lay in the house that Jack built.

This is the rat,

That ate the malt

That lay in the house that Jack built.

This is the cat,

That killed the rat,

That ate the malt

That lay in the house that Jack built.

$$p(\text{house} \mid \text{this is the}) = \frac{c(\text{this is the house})}{c(\text{this is the ...})} = \frac{1}{4}$$

Toy corpus

This is the house that Jack built.

This is the malt

That lay in the house that Jack built.

This is the rat,

That ate the malt

That lay in the house that Jack built.

This is the cat,

That killed the rat,

That ate the malt

That lay in the house that Jack built.

4-grams

$$p(\text{house} \mid \text{this is the}) = \frac{c(\text{this is the house})}{c(\text{this is the ...})} = \frac{1}{4}$$

Toy corpus

*This is the house **that** Jack built.*

This is the malt

*That lay in the house **that** Jack built.*

This is the rat,

That ate the malt

*That lay in the house **that** Jack built.*

This is the cat,

That killed the rat,

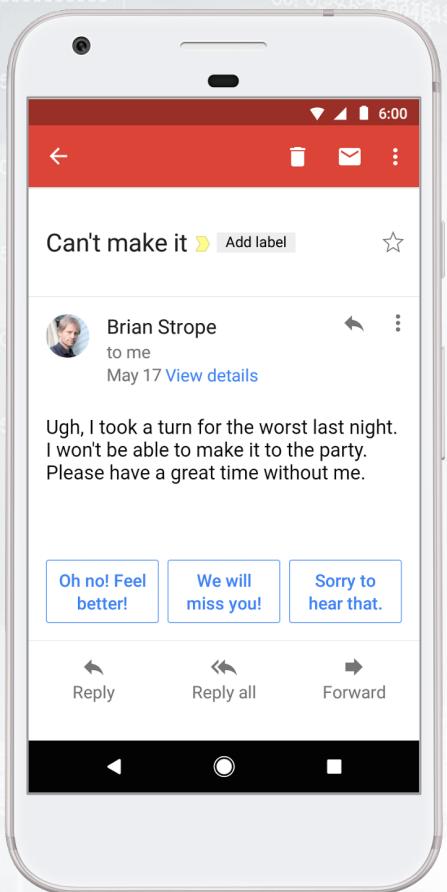
That ate the malt

*That lay in the house **that** Jack built.*

bigrams

$$p(\text{Jack} \mid \text{that}) = \frac{c(\text{that Jack})}{c(\text{that...})} = \frac{4}{10}$$

Where do we need LM?



Smart Reply / research.googleblog.com

- Suggestions in messengers
- Spelling correction
- Machine translation
- Speech recognition
- Handwriting recognition
- ...

Language modeling

This is the ...

house

rat

did

malt

What's the probability of the whole sequence?

$$p(\text{this is the house}) = ?$$

Let's do some math

Predict probability of a sequence of words:

$$\mathbf{w} = (w_1 w_2 w_3 \dots w_k)$$

Let's do some math

Predict probability of a sequence of words:

$$\mathbf{w} = (w_1 w_2 w_3 \dots w_k)$$

- **Chain rule:**

$$p(\mathbf{w}) = p(w_1)p(w_2|w_1)\dots p(w_k|w_1 \dots w_{k-1})$$

Let's do some math

Predict probability of a sequence of words:

$$\mathbf{w} = (w_1 w_2 w_3 \dots w_k)$$

- **Chain rule:**

$$p(\mathbf{w}) = p(w_1)p(w_2|w_1) \dots p(w_k | \cancel{w_1 \dots w_{k-1}})$$

- **Markov assumption:**

$$p(w_i | w_1 \dots w_{i-1}) = p(w_i | w_{\textcolor{teal}{i-n+1}} \dots w_{i-1})$$

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = p(w_1)p(w_2|w_1) \dots p(w_k|w_{k-1})$$

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = p(w_1)p(w_2|w_1) \dots p(w_k|w_{k-1})$$

Toy corpus:

This is the malt

That lay in the house that Jack built.

$$p(\text{this is the house}) = p(\text{this}) p(\text{is} | \text{this}) p(\text{the} | \text{is}) p(\text{house} | \text{the})$$

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = p(w_1)p(w_2|w_1) \dots p(w_k|w_{k-1})$$

Toy corpus:

This is the malt

That lay in the house that Jack built.

1/12

1

1

1/2

$$p(\text{this is the house}) = p(\text{this}) p(\text{is} | \text{this}) p(\text{the} | \text{is}) p(\text{house} | \text{the})$$

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = p(w_1|start) \cdot \cancel{p(w_1)} p(w_2|w_1) \dots p(w_k|w_{k-1})$$

Toy corpus:

This is the malt

That lay in the house that Jack built.

$$p(this \text{ is } the \text{ house}) = p(this) p(is|this) p(the|is) p(house|the)$$

1/2 1 1 1/2

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = p(w_1|start) \cdot p(w_2|w_1) \dots p(w_k|w_{k-1})$$

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = \cancel{p(w_1)} p(w_2|w_1) \dots p(w_k|w_{k-1}) \\ p(w_1|start)$$

It's normalized separately for each sequence length!

$$p(this) + p(that) = 1.0$$

$$p(this\ this) + p(this\ is) + \dots + p(built\ built) = 1.0$$

...

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = \cancel{p(w_1)} p(w_2|w_1) \dots p(w_k|w_{k-1}) \\ p(w_1|start) \quad \quad \quad p(end|w_k)$$

It's normalized separately for each sequence length!

$$p(this) + p(that) = 1.0$$

$$p(this\ this) + p(this\ is) + \dots + p(built\ built) = 1.0$$

...

Let's check the model

1.0000000000
0.500000
0.000000
-0.500000
-1.0000000000

▲ 60; 0.9178592142
▲ 80; 0.8915486389
◆ 60; 0.7833362060
● 30; 0.6499716055
● 30; 0.6398330960
● 30; 0.4522583897
● 120; 0.585397266
● 120; 0.3743391973

_ dog _
_ dog cat tiger _
_ cat dog cat _

$p(cat \: dog \: cat) =$

-0.115781	0.173160	0.229220	0.263536	-0.335684	0.385298
-0.115939	0.174295	-0.293325	-0.284602	-0.336713	0.386261
0.118955	0.172430	-0.231423	0.265687	-0.337730	0.387223
0.115994	0.155757	0.155234	0.610526	0.583780	0.552311
0.1152193	0.150479	0.633813	0.610331	0.583203	0.551838
0.1152003	0.050174	0.632410	0.609533	0.682515	0.550983

0.080973	0.297762	-0.554354	-0.534707	-0.510990	-0.483385
-0.002003	0.000130	0.000170	0.000217	0.000257	-0.000285
0.115994	0.155757	0.155234	0.610526	0.583780	0.552311
0.1152193	0.150479	0.633813	0.610331	0.583203	0.551838
0.1152003	0.050174	0.632410	0.609533	0.682515	0.550983

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000	360; 0.0000000000
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

Let's check the model

1.0000000000
0.5000000000
0.0000000000
-0.5000000000
-1.0000000000

△ 60; 0.9178592142
△ 80; 0.8915486389
◆ 60; 0.7833362060
● 30; 0.6499716055
● 30; 0.6398330960
● 30; 0.4322583897

_ dog _
_ dog cat tiger _
_ cat dog cat _

$$p(cat \, dog \, cat) = p(cat \mid _)$$

dog

cat

Let's check the model

1.0000000000
0.500000
0.000000
-0.500000
-1.0000000000

△ 60; 0.9178592142
△ 80; 0.8915486389
◆ 60; 0.7833362060
● 30; 0.6499716055
● 30; 0.6398330960
● 30; 0.4522583697
● 120; 0.5853979766
● 120; 0.3743391973

_ dog _

_ dog cat tiger _

_ cat dog cat _

$$p(cat \, dog \, cat) = p(cat \mid _)$$

dog

cat

Let's check the model

1.0000000000
0.500000
0.000000
-0.500000
-1.0000000000

dog _
_ dog cat tiger _
_ cat dog cat _

$$p(cat \, dog \, cat) = p(cat \mid _) \, p(dog \mid cat)$$

dog

cat tiger

cat dog

cat _

Let's check the model

_ dog _
_ dog cat tiger _
_ cat dog cat _

$$p(cat \, dog \, cat) = p(cat \mid _) \, p(dog \mid cat)$$

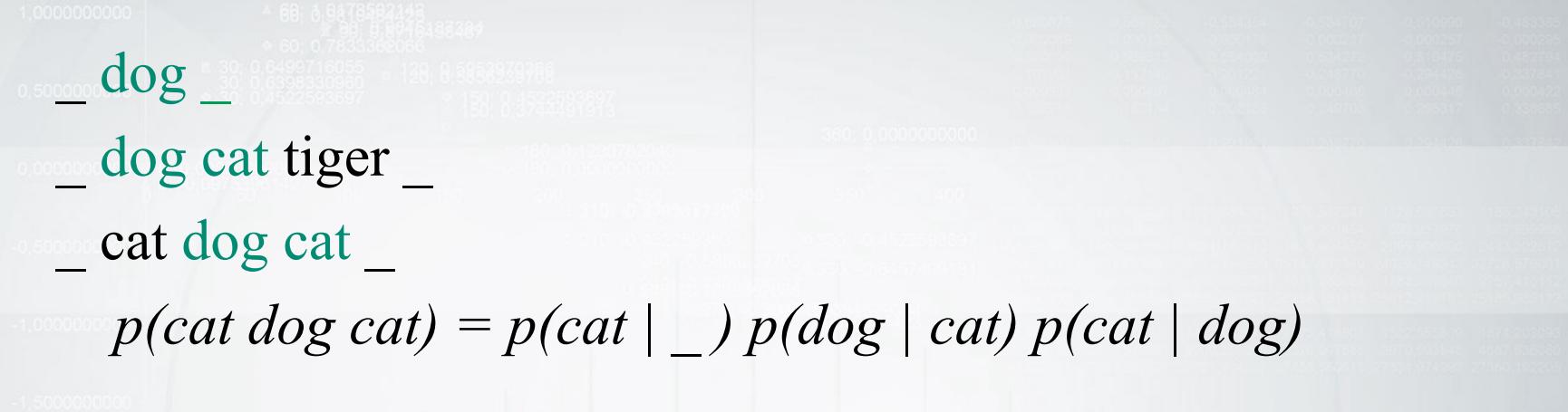
dog

cat tiger

cat dog

cat _

Let's check the model



dog

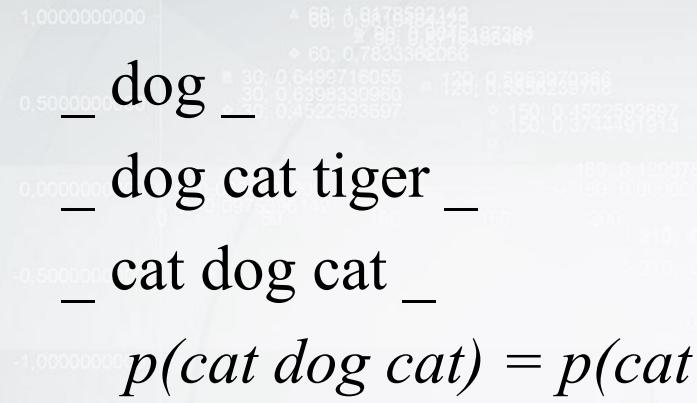
cat tiger

cat dog cat

cat dog_

cat _

Let's check the model



$$p(cat\ dog\ cat) = p(cat | _) p(dog | cat) p(cat | dog)$$

dog

cat tiger

cat dog cat

cat dog_

cat _

Let's check the model

_ dog _
_ dog cat tiger _
_ cat dog cat _

$$p(cat \, dog \, cat) = p(cat \mid _) \, p(dog \mid cat) \, p(cat \mid dog) \, p(_ \mid cat)$$

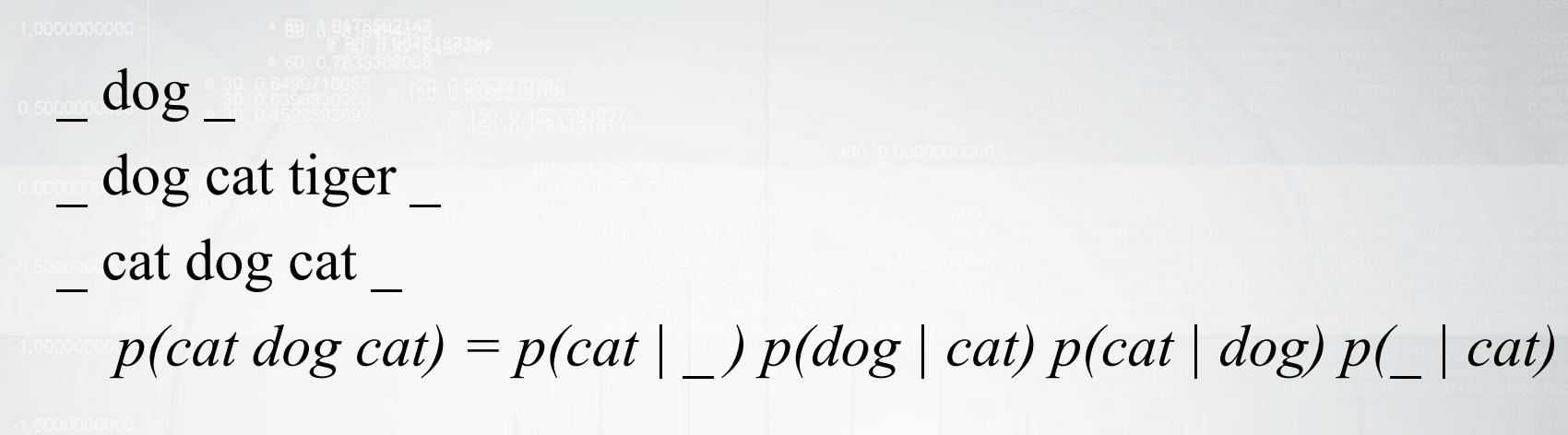
dog

cat tiger

cat dog cat tiger cat dog_
cat dog cat dog
cat dog cat _

cat _

Let's check the model



dog

cat tiger

cat dog cat tiger

cat dog cat dog

cat dog cat _

cat _

Resume: bigram language model

Define the model:

$$p(\mathbf{w}) = \prod_{i=1}^{k+1} p(w_i | w_{i-1})$$

Estimate the probabilities:

$$p(w_i | w_{i-1}) = \frac{c(w_{i-1} w_i)}{\sum_{\mathbf{w}_i} c(w_{i-1} \mathbf{w}_i)} = \frac{c(w_{i-1} w_i)}{c(w_{i-1})}$$

It's all about counting!