13. oktober 2022

IN1140 Gruppe 3

Denne økta

Summering og multiplisering

Betinget sannsynlighet

n-gram

Trigrammodell

Summering og multiplisering

De greske bokstavene $\boldsymbol{\Sigma}$ og $\boldsymbol{\Pi}$ brukes hhv. når vi systematisk skal summere og multiplisere.

$$\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5 = 15$$

$$\prod_{i=1}^{5} i = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

$$\sum_{i=2}^{6} \frac{i}{2} = \frac{2}{2} + \frac{3}{2} + \frac{4}{2} + \frac{5}{2} + \frac{6}{2} = \frac{20}{2} = 10$$

$$\prod_{i=2}^{6} \frac{i}{2} = \frac{2}{2} \cdot \frac{3}{2} \cdot \frac{4}{2} \cdot \frac{5}{2} \cdot \frac{6}{2} = \frac{720}{32} = 22{,}5$$

Oppgave

Regn ut følgende:

$$\sum_{i=2}^{4} i^2$$

$$\prod_{i=1}^{2} 3^i$$

$$\sum_{i=6}^{7} i + 2i$$

$$\prod_{i=4}^{7} \frac{2}{i}$$

Løsning

$$\sum_{i=2}^{4} i^2 = 2^2 + 3^2 + 4^2 = 29$$

$$\prod_{i=1}^{2} 3^i = 3^1 \cdot 3^2 = 27$$

$$\sum_{i=6}^{7} i + 2i = (6 + 2 \cdot 6) + (7 + 2 \cdot 7) = 39$$

$$\prod_{i=4}^{7} \frac{2}{i} = \frac{2}{4} \cdot \frac{2}{5} \cdot \frac{2}{6} \cdot \frac{2}{7} = \frac{16}{840} = \frac{2}{105}$$

Betinget sannsynlighet

w_n er orde nummer n. w_{n-1} er ordet foran.
$count()$ betyr bare at vi må telle det som står inni parentesene (argumentet).

Fra Jurafsky & Martin: $P(w_n|w_{n-1}) = \frac{count(w_{n-1}\ w_n)}{count(w_{n-1})}$

Dette kan brukes til å beregne sannsynligheten til en ordsekvens, der k er antall ord i sekvensen:

$$P(w_1 \ldots w_k) = \prod_{i=1}^{k} P(w_i|w_{i-1})$$

n-gram

Tokenisert setning: `Mons satt fast i treet .`

Unigram ($n = 1$):
`(Mons), (satt), (fast), (i), (treet), (.)`

Bigram ($n = 2$):
`(Mons, satt), (satt, fast), (fast, i), (i, treet,), (treet, .)`

Bigram ($n = 3$):
`(Mons, satt, fast), (satt, fast, i), (fast, i, treet), (i, treet, .)`

Sannsynlighet for setning med bigram

Gitt korpus:
```
<s> Lars trener aldri <\s>
<s> Tor løper sjeldent <\s>
<s> Tor trener ofte <\s>
```

Beregne sannsynlighet for: `<s> Tor trener aldri <\s>`

Vi behøver følgende bigram-sannsynligheter:

$$\mathrm{P}(Tor|<s>) = \frac{count(<s>\,Tor)}{count(<s>)} = \frac{2}{3} \qquad \mathrm{P}(trener|Tor) = \frac{count(Tor\ trener)}{count(Tor)} = \frac{1}{2}$$

$$\mathrm{P}(aldri|trener) = \frac{count(trener\ aldri)}{count(trener)} = \frac{1}{2} \qquad \mathrm{P}(<\mathrm{s}>|aldri) = \frac{count(aldri\ <\mathrm{s}>)}{count(aldri)} = \frac{1}{1}$$

Sannsynlighet for setning med bigram

$$\mathrm{P}(Tor|<s>) = \frac{count(<s>\,Tor)}{count(<s>)} = \frac{2}{3} \qquad \mathrm{P}(trener|Tor) = \frac{count(Tor\ trener)}{count(Tor)} = \frac{1}{2}$$

$$\mathrm{P}(aldri|trener) = \frac{count(trener\ aldri)}{count(trener)} = \frac{1}{2} \qquad \mathrm{P}(<\mathrm{s}>|aldri) = \frac{count(aldri\ <\mathrm{s}>)}{count(aldri)} = \frac{1}{1}$$

$$P(<\mathrm{s}>\ Tor\ \ trener\ \ aldri\ <\mathrm{s}>)$$

$$= \mathrm{P}(Tor|<s>) \cdot \mathrm{P}(trener|Tor) \cdot \mathrm{P}(aldri|trener) \cdot \mathrm{P}(<\mathrm{s}>|aldri) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{2}{12} = \frac{1}{6}$$