

Gruppetime #6

IN1140, gruppe 1

Betinget sannsynlighet

w_n er et hvilket som helst ord, og w_{n-1} er ordet foran.

$\text{count}()$ betyr helt enkelt at vi må telle det som står inne i parenteser (som argument).

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

Kan brukes til å beregne sannsynligheten til en ordsekvens:

k er antall ord i sekvensen.

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

n-grammer

Setning:

Mons satt fast i treet .

Unigrammer (n = 1):

`<Mons>, <satt>, <fast>, <i>, <treet>, <.>`

Bigrammer (n = 2):

`<Mons, satt>, <satt, fast>, <fast, i>, <i, treet>, <treet, .>`

Trigrammer (n = 3):

`<Mons, satt, fast>, <satt, fast, i>, <fast, i, treet>, <i, treet, .>`

Sannsynlighet for setning med bigrammer

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>
... altså $P(\langle s \rangle \text{ Tor trener aldri } \langle /s \rangle)$

Bigram-sannsynligheter vi trenger da:

$$P(\text{Tor} | \langle s \rangle) = \frac{\text{count}(\langle s \rangle \text{ Tor})}{\text{count}(\langle s \rangle)} = \frac{2}{3}$$

$$P(\text{trener} | \text{Tor}) = \frac{\text{count}(\text{Tor trener})}{\text{count}(\text{Tor})} = \frac{1}{2}$$

$$P(\text{aldri} | \text{trener}) = \frac{\text{count}(\text{trener aldri})}{\text{count}(\text{trener})} = \frac{1}{2}$$

$$P(\langle /s \rangle | \text{aldri}) = \frac{\text{count}(\text{aldri } \langle /s \rangle)}{\text{count}(\text{aldri})} = \frac{1}{1}$$

Utregning

$$P(\text{Tor} | \langle s \rangle) = \frac{\text{count}(\langle s \rangle \text{Tor})}{\text{count}(\langle s \rangle)} = \frac{2}{3}$$

$$P(\text{trener} | \text{Tor}) = \frac{\text{count}(\text{Tor trener})}{\text{count}(\text{Tor})} = \frac{1}{2}$$

$$P(\text{aldri} | \text{trener}) = \frac{\text{count}(\text{trener aldri})}{\text{count}(\text{trener})} = \frac{1}{2}$$

$$P(\langle \backslash s \rangle | \text{ofte}) = \frac{\text{count}(\text{aldri} \langle \backslash s \rangle)}{\text{count}(\text{aldri})} = \frac{1}{1}$$

$$P(\langle s \rangle \text{Tor trener aldri} \langle \backslash s \rangle)$$

$$= P(\text{Tor} | \langle s \rangle) * P(\text{trener} | \text{Tor}) * P(\text{aldri} | \text{trener}) P(\langle \backslash s \rangle | \text{aldri}) = \frac{2}{3} * \frac{1}{2} * \frac{1}{2} * \frac{1}{1} = \frac{2}{12} = \frac{1}{6}$$