Naive Bayes: Utregningseksempel

liljacs@uio.no
IN1140

"Jeg liker brus"

Gitt korpuset vårt – er setningen mest sannsynlig positiv eller negativ?

KORPUS

POS	Jeg elsker brus
POS	Jeg liker kake
NEG	Jeg hater brus

- Vi må først ha prior-sannsynligheten for hver klasse:

$P(POS) = \frac{2}{3}$ ← 2 dokumenter i den Positive klassen

av totalt 3 dokumenter

$P(NEG) = \frac{1}{3}$

- Så må vi regne ut likelihood-sannsynlighetene. Siden ikke alle ordene finnes i hver klasse, legger vi på "add-one-smoothing". Da ser formelen slik ut:

$$P(w_i|c) = \frac{\text{Antall forekomster av ordet i klassen} + 1}{\text{Antall ord i hele klassen} + |V|}$$

$|V|$ = antall unike ord i hele treningssettet

- Hva er sannsynligheten for at setningen tilhører den positive klassen?
- Vi bruker formelen over:

$$P(Jeg|POS) = \frac{2+1}{6+6}, \quad P(liker|POS) = \frac{1+1}{6+6}, \quad P(brus|POS) = \frac{1+1}{6+6}$$

Så fullfører vi formelen ved å gange dette sammen, i tillegg til prior-sannsynligheten:

$$\frac{2}{3} \cdot \frac{3}{12} \cdot \frac{2}{12} \cdot \frac{2}{12} \approx 0.462\%$$

↓

Så gjør vi det samme for den negative klassen:

$$P(\text{Jeg}|\text{NEG}) = \frac{1+1}{3+6}, \quad P(\text{liker}|\text{NEG}) = \frac{0+1}{3+6}, \quad P(\text{brus}|\text{NEG}) = \frac{1+1}{3+6}$$

$$\frac{1}{3} \cdot \frac{1+1}{3+6} \cdot \frac{0+1}{3+6} \cdot \frac{1+1}{3+6} \approx 0.182\%$$

I følge korpuset vårt, er det størst sannsynlighet for at setningen tilhører klassen positiv.