

# NAIVE BAYES 🧠💥


En probabilistisk modell som har som mål å predikere den mest sannsynlige klassen for et dokument

# VEILEDET LÆRING

## TRENING

- Annotert treningsdata
  - Treningseksempler med en «merkelapp»- en fasit
  - Eks: dokumenter annotert med tema, f.eks. «sport»
- Vi trener klassifikatoren med treningsdataen, slik at den kan anvendes på nye og *usette* data
- Jo mer treningsdata, desto bedre modell

## EVALUERING

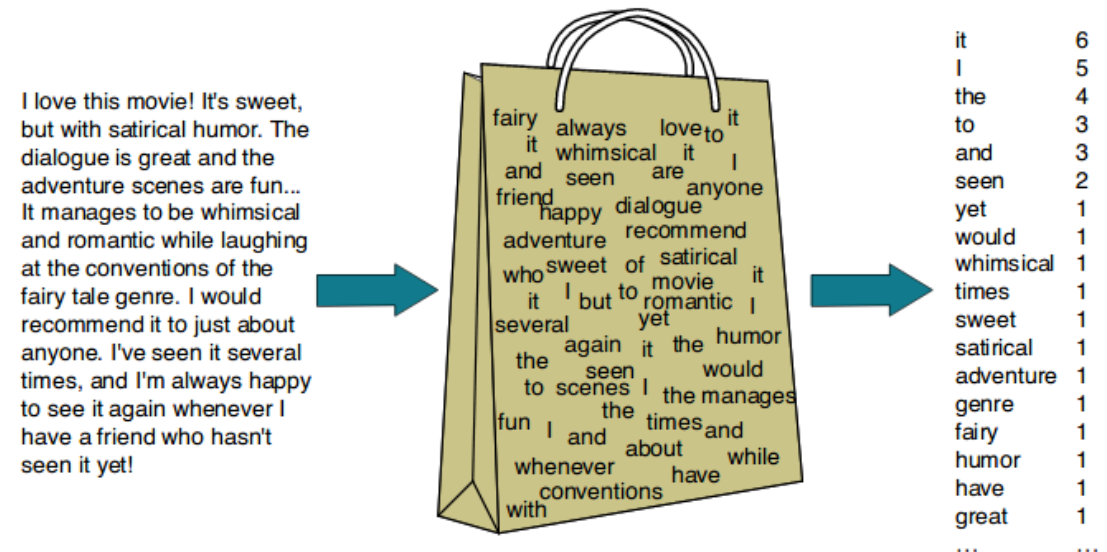
- Vi må også *teste* med annoterte data
- Sammenligner modellens prediksjoner med «gullstandarden» 
- $Accuracy = \frac{\text{korrekte prediksjoner}}{\text{totalt antall prediksjoner}}$

# TREKK-REPRESENTASJON

- Første skritt i å trene en modell er å velge hvordan vi skal representere dataene
- Vi representerer dataene med *trekk*

## BAG-OF-WORDS

- Vanligst for tekstklassifikasjon
- Hvert ord i dokumentet er et trekk
- Rekkefølge eller posisjon har ikke noe å si



# 1. HOVEDANTAGELSE

$$\hat{C} = \underset{c \in C}{\operatorname{argmax}} P(c|d)$$

Velg den mest sannsynlige klassen

Predikert klasse

hver klasse  $c$  i samlingen av alle klasser  $C$

sannsynligheten for en klasse gitt et dokument

# 2. OMSKRIVING MED BAYES TEOREM

Betinget sannsynlighet + produktsetningen

$$\frac{P(d|c)P(c)}{\cancel{P(d)}} = P(d|c)P(c)$$

$P(d)$  er uavhengig av  $c$ , og vil være konstant for hver  $d$

ETTER OMSKRIVING:

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(d|c) P(c)$$

### 3. **NAIV** UAVHENGIGHETSANTAGELSE

Vi erstatter  $P(d|c)$  med multiplikasjonsregelen for uavhengige hendelser, siden vi **naivt** antar at hvert trekk er uavhengig

$$P(c|d) = \underbrace{P(w_1, w_2, \dots, w_k | c) P(c)}_{\text{Produktet av alle sannsynlighetene for hvert ord, gitt en klasse}} \approx \prod_{i=1}^k P(w_i | c) P(c)$$

# ENDELIG FORMEL

$$\hat{c} = \operatorname{argmax} P(c) \prod_{w \in d} P(w | c)$$

Prior-sannsynlighet for klassen  $c$

Likelihood-sannsynlighet

Multipliser sammen alle sannsynlighetene for hvert ord gitt en klasse i dokumentet

## PRIOR-SANNSYNLIGHET

$$P(c) = \frac{\textit{Antall dokumenter med klassen}}{\textit{Totalt antall dokumenter}}$$

## LIKELIHOOD-SANNSYNLIGHET

$$P(w_i|c) = \frac{\textit{Antall forekomster av ordet i klassen}}{\textit{Antall forekomster av alle ord i klassen}}$$

# SMOOTHING

Hva om et ord ikke forekommer i en klasse? 🤔

- Løsningen er *smoothing*!

*Add-one-smoothing/Laplace smoothing:*

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