Naive Bayes vs. Maxent models

Generative vs. Discriminative models: The problem of overcounting evidence

Christopher Manning

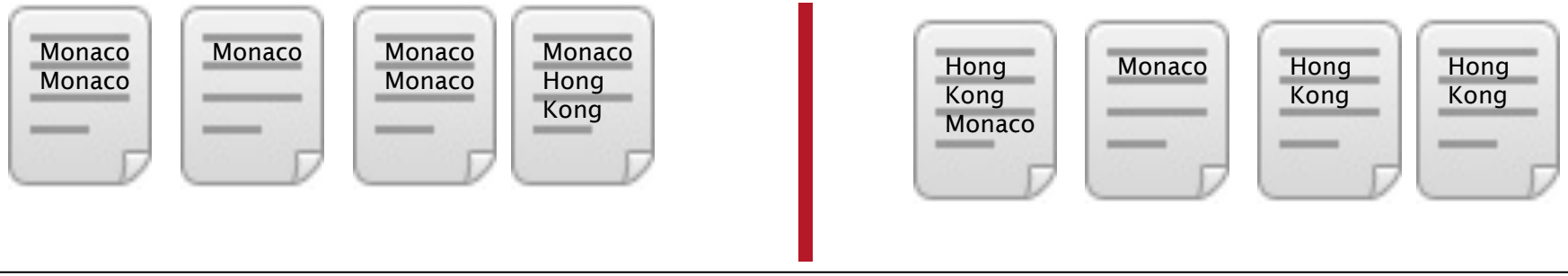
Christopher Manning

Text classification: Asia or Europe

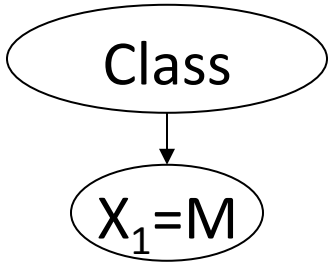
Europe

Training Data

Asia



NB Model



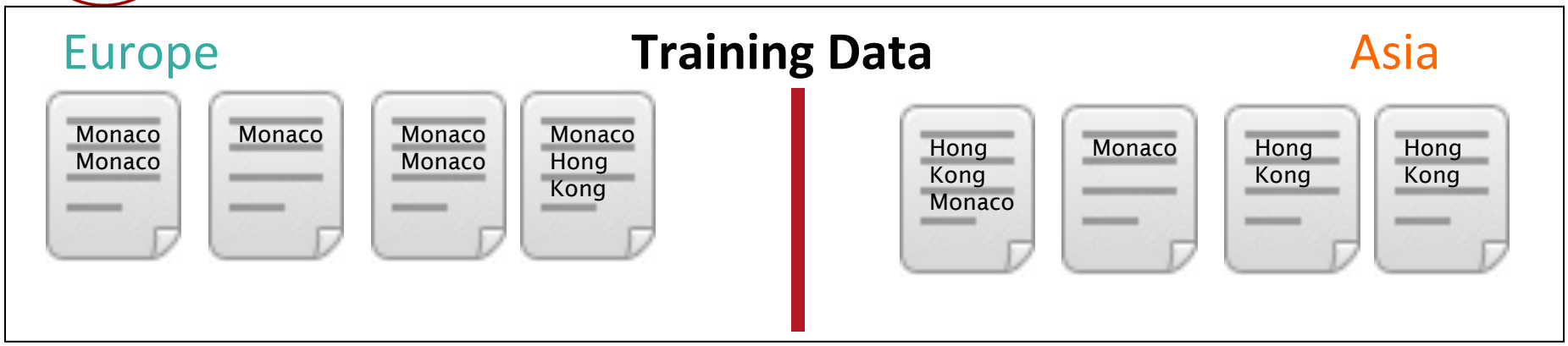
NB FACTORS:

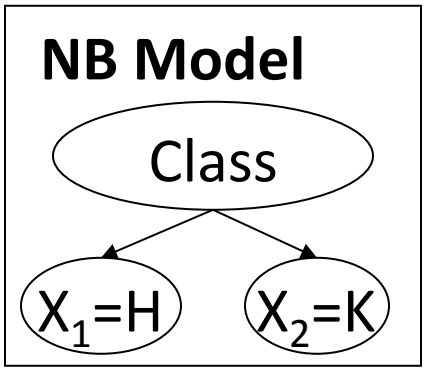
- P(A) = P(E) =
- P(M|A) =
- P(M|E) =

PREDICTIONS:

- P(A,M) =
- P(E,M) =
- P(A|M) =
- P(E|M) =

Text classification: Asia or Europe





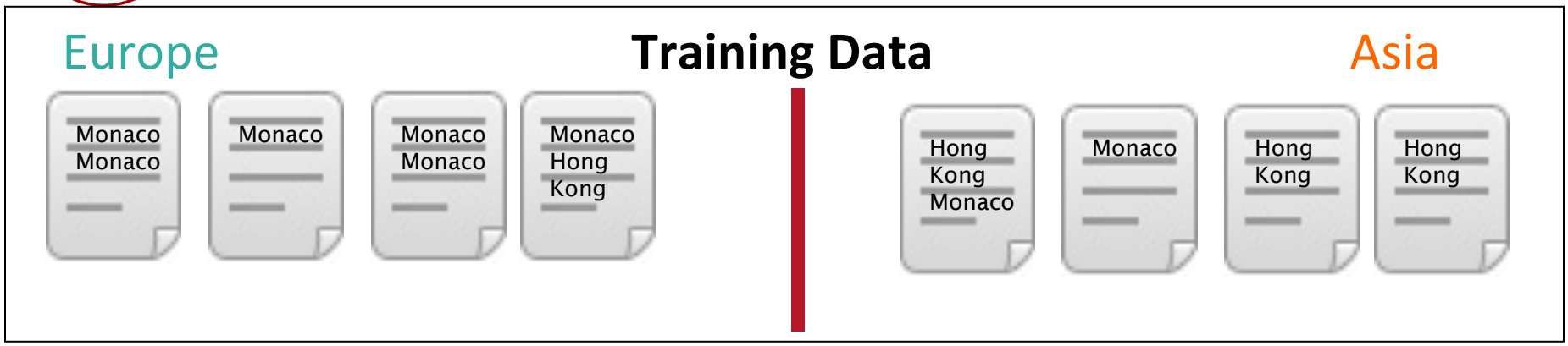
NB FACTORS:

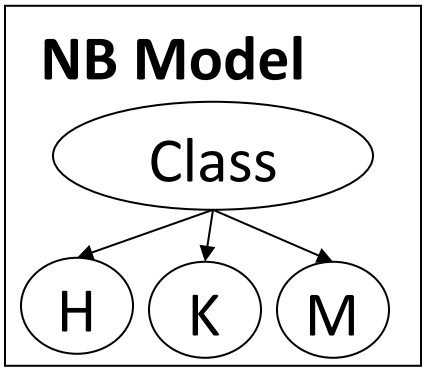
- P(A) = P(E) =
- P(H|A) = P(K|A) =
- P(H|E) = PK|E) =

PREDICTIONS:

- P(A,H,K) =
- P(E,H,K) =
- P(A|H,K) =
- P(E|H,K) =

Text classification: Asia or Europe





NB FACTORS:

- P(A) = P(E) =
- P(M|A) =
- P(M|E) =
- P(H|A) = P(K|A) =
- P(H|E) = PK|E) =

PREDICTIONS:

- P(A,H,K,M) =
- P(E,H,K,M) =
- P(A|H,K,M) =
- P(E|H,K,M) =

Naive Bayes vs. Maxent Models

- Naive Bayes models multi-count correlated evidence
  - Each feature is multiplied in, even when you have multiple features telling you the same thing
- Maximum Entropy models (pretty much) solve this problem
  - As we will see, this is done by weighting features so that model expectations match the observed (empirical) expectations

Naive Bayes vs. Maxent models

Generative vs. Discriminative models: The problem of overcounting evidence

Christopher Manning