Learning Probabilistic Models

Dr. Steven Bethard

Computer and Information Sciences
University of Alabama at Birmingham

14 Apr 2016

Outline

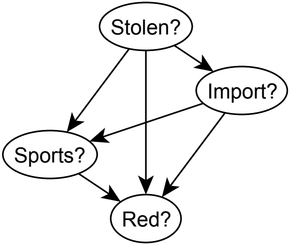
1. Naive Bayes
   - Models
   - Examples
   - Properties
2. Expectation Maximization
   - Hidden Variables
   - Algorithm
   - Properties

Outline

Naive Bayes Networks

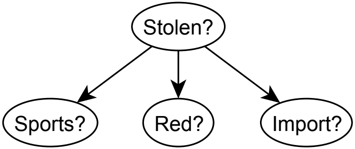
Bayesian Network

Represents all variable dependence relations



Naive Bayes Network

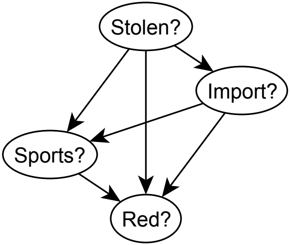
Assumes features are conditionally independent given the class variable



Naive Bayes Networks

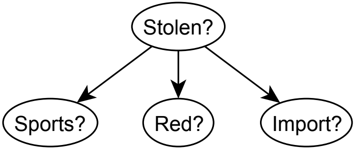
Bayesian Network

Represents all variable dependence relations



Naive Bayes Network

Assumes features are conditionally independent given the class variable



Naive Bayes Models

Given class C and features $F_1, \ldots, F_n$

$$\mathbf{P}(C|F_1, F_2, \ldots, F_n)$$
$$= \alpha \mathbf{P}(F_1, F_2, \ldots, F_n|C)\mathbf{P}(C) \quad \text{Bayes' Rule}$$
$$= \alpha \mathbf{P}(F_1|C)\mathbf{P}(F_2|C) \ldots \mathbf{P}(F_n|C)\mathbf{P}(C) \quad \text{Naive Bayes}$$

Naive Bayes models

$$\mathbf{P}(C|F_1, F_2, \ldots, F_n) = \alpha \mathbf{P}(C) \prod_{i=1}^{n} \mathbf{P}(F_i|C)$$

Training models

- Find the probability of each class
- Find the probability of each feature given the class

Naive Bayes Models

Given class C and features $F_1, \ldots, F_n$

$$\begin{aligned} &\mathbf{P}(C|F_1, F_2, \ldots, F_n) \\ &\quad = \alpha \mathbf{P}(F_1, F_2, \ldots, F_n|C)\mathbf{P}(C) && \text{Bayes' Rule} \\ &\quad = \alpha \mathbf{P}(F_1|C)\mathbf{P}(F_2|C)\ldots\mathbf{P}(F_n|C)\mathbf{P}(C) && \text{Naive Bayes} \end{aligned}$$

Naive Bayes models

$$\mathbf{P}(C|F_1, F_2, \ldots, F_n) = \alpha \mathbf{P}(C) \prod_{i=1}^{n} \mathbf{P}(F_i|C)$$

Training models

- Find the probability of each class
- Find the probability of each feature given the class

Naive Bayes Models

Given class C and features $F_1, \ldots, F_n$

$$\begin{aligned}
&\mathbf{P}(C|F_1, F_2, \ldots, F_n) \\
&\quad = \alpha \mathbf{P}(F_1, F_2, \ldots, F_n|C)\mathbf{P}(C) && \text{Bayes' Rule} \\
&\quad = \alpha \mathbf{P}(F_1|C)\mathbf{P}(F_2|C)\ldots\mathbf{P}(F_n|C)\mathbf{P}(C) && \text{Naive Bayes}
\end{aligned}$$

Naive Bayes models

$$\mathbf{P}(C|F_1, F_2, \ldots, F_n) = \alpha \mathbf{P}(C) \prod_{i=1}^{n} \mathbf{P}(F_i|C)$$

Training models

- Find the probability of each class
- Find the probability of each feature given the class

Naive Bayes Models

Given class C and features $F_1, \ldots, F_n$

$$\begin{aligned} & \mathbf{P}(C|F_1, F_2, \ldots, F_n) \\ &\quad = \alpha \mathbf{P}(F_1, F_2, \ldots, F_n|C)\mathbf{P}(C) && \text{Bayes' Rule} \\ &\quad = \alpha \mathbf{P}(F_1|C)\mathbf{P}(F_2|C) \ldots \mathbf{P}(F_n|C)\mathbf{P}(C) && \text{Naive Bayes} \end{aligned}$$

Naive Bayes models

$$\mathbf{P}(C|F_1, F_2, \ldots, F_n) = \alpha \mathbf{P}(C) \prod_{i=1}^{n} \mathbf{P}(F_i|C)$$

Training models

- Find the probability of each class
- Find the probability of each feature given the class

Naive Bayes Models

Given class C and features $F_1, \ldots, F_n$

$$\begin{aligned}
&\mathbf{P}(C|F_1, F_2, \ldots, F_n) \\
&\quad = \alpha \mathbf{P}(F_1, F_2, \ldots, F_n|C)\mathbf{P}(C) && \text{Bayes' Rule} \\
&\quad = \alpha \mathbf{P}(F_1|C)\mathbf{P}(F_2|C) \ldots \mathbf{P}(F_n|C)\mathbf{P}(C) && \text{Naive Bayes}
\end{aligned}$$

Naive Bayes models

$$\mathbf{P}(C|F_1, F_2, \ldots, F_n) = \alpha \mathbf{P}(C) \prod_{i=1}^{n} \mathbf{P}(F_i|C)$$

Training models

- Find the probability of each class
- Find the probability of each feature given the class

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$
$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes   At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$

$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$
$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$
$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$
$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16}+\frac{1}{24}}$

Naive Bayes Classification

Origin?	Color?	Type?	Stolen?
import	red	sports	yes
import	red	sports	yes
domestic	white	sports	yes
domestic	red	van	yes
import	red	van	no
domestic	white	sports	no

A domestic red sports car

$$P(y|d,r,s) = \alpha P(y)P(d|y)P(r|y)P(s|y) = \alpha \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{3}{16}\alpha$$

$$P(n|d,r,s) = \alpha P(n)P(d|n)P(r|n)P(s|n) = \alpha \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{24}\alpha$$

Predict stolen? yes

At what probability? $\frac{9}{11} = \frac{\frac{3}{16}}{\frac{3}{16} + \frac{1}{24}}$

Naive Bayes Exercise

Ends with -ed	Initial Capital	Part of Speech
no	no	noun
no	yes	noun
no	yes	noun
yes	no	noun
yes	yes	noun
no	no	verb
yes	no	verb
yes	yes	verb

Assign part of speech tags: John tripped

Naive Bayes Exercise

Ends with -ed	Initial Capital	Part of Speech
no	no	noun
no	yes	noun
no	yes	noun
yes	no	noun
yes	yes	noun
no	no	verb
yes	no	verb
yes	yes	verb

Assign part of speech tags: John tripped

	John	tripped
	noun	verb
	$\frac{27}{32}$	$\frac{5}{8}$

Naive Bayes Properties

Naive Bayes assumption is hardly ever true

- Probability estimates of Naive Bayes are poor
- Classification decisions are often surprisingly good

Empirical Observations

- Works best when many equally important features
- Somewhat robust to noise (uninformative) features
- Training and classification are typically fast

Naive Bayes Properties

Naive Bayes assumption is hardly ever true

- Probability estimates of Naive Bayes are poor
- Classification decisions are often surprisingly good

Empirical Observations

- Works best when many equally important features
- Somewhat robust to noise (uninformative) features
- Training and classification are typically fast

Outline

A Coin Example

Given two coins:

A with $P(A{=}H)$ for heads and $P(A{=}T)$ for tails

B with $P(B{=}H)$ for heads and $P(B{=}T)$ for tails

Estimate $\mathbf{P}(A)$ and $\mathbf{P}(B)$ from the coin tosses:

A	HHHHTHHHHH
B	HTHTTTHHTT
A	HTHHHHHTHH
A	THHHTHHHTH
B	HTTTHHTHTH

A Coin Example

Given two coins:

A with $P(A{=}H)$ for heads and $P(A{=}T)$ for tails

B with $P(B{=}H)$ for heads and $P(B{=}T)$ for tails

Estimate $\mathbf{P}(A)$ and $\mathbf{P}(B)$ from the coin tosses:

		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
A	HHHHTHHHHH	9	1		
B	HTHTTTHHTT			4	6
A	HTHHHHHTHH	8	2		
A	THHHTHHHTH	7	3		
B	HTTTHHTHTH			5	5

A Coin Example

Given two coins:

A with $P(A=H)$ for heads and $P(A=T)$ for tails

B with $P(B=H)$ for heads and $P(B=T)$ for tails

Estimate $\mathbf{P}(A)$ and $\mathbf{P}(B)$ from the coin tosses:

		$A=H$	$A=T$	$B=H$	$B=T$
A	HHHHTHHHHH	9	1		
B	HTHTTTHHTT			4	6
A	HTHHHHHTHH	8	2		
A	THHHTHHHTH	7	3		
B	HTTTHHTHTH			5	5

$$\mathbf{P}(A)=\left\langle \frac{24}{24+6}, \frac{6}{24+6} \right\rangle \quad \mathbf{P}(B)=\left\langle \frac{9}{9+11}, \frac{11}{9+11} \right\rangle$$

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C=A)$ and $P(C=B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

		$A=H$	$A=T$	$B=H$	$B=T$
?	HHHHTHHHHH				
?	HTHTTTHHTT				
?	HTHHHHHTHH				
?	THHHTHHHTH				
?	HTTTHHTHTH				

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C=A)$ and $P(C=B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

		$A=H$	$A=T$	$B=H$	$B=T$
?	HHHHTHHHHH				
?	HTHTTTHHTT				
?	HTHHHHHTHH				
?	THHHTHHHTH				
?	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH				
0.35	0.65	HTHTTTHHTT				
0.73	0.27	HTHHHHHTHH				
0.65	0.35	THHHTHHHTH				
0.45	0.55	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C=A)$ and $P(C=B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C=A$	$C=B$		$A=H$	$A=T$	$B=H$	$B=T$
0.80	0.20	HHHHTHHHHH	7.2			
0.35	0.65	HTHTTTHHTT				
0.73	0.27	HTHHHHHTHH				
0.65	0.35	THHHTHHHTH				
0.45	0.55	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH	7.2	0.8		
0.35	0.65	HTHTTTHHTT				
0.73	0.27	HTHHHHHTHH				
0.65	0.35	THHHTHHHTH				
0.45	0.55	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH	7.2	0.8	1.8	0.2
0.35	0.65	HTHTTTHHTT				
0.73	0.27	HTHHHHHTHH				
0.65	0.35	THHHTHHHTH				
0.45	0.55	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH	7.2	0.8	1.8	0.2
0.35	0.65	HTHTTTHHTT	1.4	2.1		
0.73	0.27	HTHHHHHTHH				
0.65	0.35	THHHTHHHTH				
0.45	0.55	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH	7.2	0.8	1.8	0.2
0.35	0.65	HTHTTTHHTT	1.4	2.1	2.6	3.9
0.73	0.27	HTHHHHHTHH				
0.65	0.35	THHHTHHHTH				
0.45	0.55	HTTTHHTHTH				

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH	7.2	0.8	1.8	0.2
0.35	0.65	HTHTTTHHTT	1.4	2.1	2.6	3.9
0.73	0.27	HTHHHHHTHH	5.8	1.5	2.2	0.5
0.65	0.35	THHHTHHHTH	4.5	1.9	2.5	1.1
0.45	0.55	HTTTHHTHTH	2.2	2.2	2.8	2.8

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

A Coin Example with a Hidden Variable

But what if we didn't know which coin was being tossed?

- i.e. $P(C{=}A)$ and $P(C{=}B)$ are unknown

Estimate $\mathbf{P}(A)$, $\mathbf{P}(B)$ from the coin tosses:

$C{=}A$	$C{=}B$		$A{=}H$	$A{=}T$	$B{=}H$	$B{=}T$
0.80	0.20	HHHHTHHHHH	7.2	0.8	1.8	0.2
0.35	0.65	HTHTTTHHTT	1.4	2.1	2.6	3.9
0.73	0.27	HTHHHHHTHH	5.8	1.5	2.2	0.5
0.65	0.35	THHHTHHHTH	4.5	1.9	2.5	1.1
0.45	0.55	HTTTHHTHTH	2.2	2.2	2.8	2.8

What if we knew $\mathbf{P}(C|\text{HHHHTHHHHH})$, etc.?

$$\mathbf{P}(A) = \left\langle \frac{21.3}{21.3+8.6}, \frac{8.6}{21.3+8.6} \right\rangle \quad \mathbf{P}(B) = \left\langle \frac{11.7}{11.7+8.4}, \frac{8.4}{11.7+8.4} \right\rangle$$

Probability of the Hidden Variable

$P(C{=}A|\text{HHHHTHHHHH})$

$={\alpha}P(\text{HHHHTHHHHH}|C{=}A)P(C{=}A)$

$={\alpha}P(\text{H}|C{=}A)P(\text{H}|C{=}A)\ldots P(\text{H}|C{=}A)P(C{=}A)$

$={\alpha}P(\text{H}|A)P(\text{H}|A)\ldots P(\text{H}|A)P(C{=}A)$

Say we make an initial guess:

- $\mathbf{P}(C){=}\langle 0.5, 0.5\rangle$
- $\mathbf{P}(A){=}\langle 0.6, 0.4\rangle$
- $\mathbf{P}(B){=}\langle 0.5, 0.5\rangle$

Then:

$P(C{=}A|\text{HHHHTHHHHH}){=}\alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5{\approx}.00202\alpha$

$P(C{=}B|\text{HHHHTHHHHH}){=}\alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5{\approx}.00049\alpha$

$\mathbf{P}(C|\text{HHHHTHHHHH}){\approx}\langle .80, .20\rangle$

Probability of the Hidden Variable

$P(C{=}A|\mathrm{HHHHTHHHHH})$
$=\alpha P(\mathrm{HHHHTHHHHH}|C{=}A)P(C{=}A)$
$=\alpha P(\mathrm{H}|C{=}A)P(\mathrm{H}|C{=}A)\ldots P(\mathrm{H}|C{=}A)P(C{=}A)$
$=\alpha P(\mathrm{H}|A)P(\mathrm{H}|A)\ldots P(\mathrm{H}|A)P(C{=}A)$

Say we make an initial guess:

- $\mathbf{P}(C){=}\langle 0.5, 0.5\rangle$
- $\mathbf{P}(A){=}\langle 0.6, 0.4\rangle$
- $\mathbf{P}(B){=}\langle 0.5, 0.5\rangle$

Then:

$P(C{=}A|\mathrm{HHHHTHHHHH}){=}\alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5{\approx}.00202\alpha$
$P(C{=}B|\mathrm{HHHHTHHHHH}){=}\alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5{\approx}.00049\alpha$
$\mathbf{P}(C|\mathrm{HHHHTHHHHH}){\approx}\langle .80, .20\rangle$

Probability of the Hidden Variable

$P(C{=}A|\mathrm{HHHHTHHHHH})$
$=\alpha P(\mathrm{HHHHTHHHHH}|C{=}A)P(C{=}A)$
$=\alpha P(\mathrm{H}|C{=}A)P(\mathrm{H}|C{=}A)\ldots P(\mathrm{H}|C{=}A)P(C{=}A)$
$=\alpha P(\mathrm{H}|A)P(\mathrm{H}|A)\ldots P(\mathrm{H}|A)P(C{=}A)$

Say we make an initial guess:

- $\mathbf{P}(C){=}\langle 0.5, 0.5\rangle$
- $\mathbf{P}(A){=}\langle 0.6, 0.4\rangle$
- $\mathbf{P}(B){=}\langle 0.5, 0.5\rangle$

Then:

$P(C{=}A|\mathrm{HHHHTHHHHH}){=}\alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5{\approx}.00202\alpha$
$P(C{=}B|\mathrm{HHHHTHHHHH}){=}\alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5{\approx}.00049\alpha$
$\mathbf{P}(C|\mathrm{HHHHTHHHHH}){\approx}\langle .80, .20\rangle$

Probability of the Hidden Variable

$P(C{=}A|\text{HHHHTHHHHH})$
$={\alpha}P(\text{HHHHTHHHHH}|C{=}A)P(C{=}A)$
$={\alpha}P(\texttt{H}|C{=}A)P(\texttt{H}|C{=}A)\dots P(\texttt{H}|C{=}A)P(C{=}A)$
$={\alpha}P(\texttt{H}|A)P(\texttt{H}|A)\dots P(\texttt{H}|A)P(C{=}A)$

Say we make an initial guess:

- $\mathbf{P}(C){=}\langle 0.5, 0.5\rangle$
- $\mathbf{P}(A){=}\langle 0.6, 0.4\rangle$
- $\mathbf{P}(B){=}\langle 0.5, 0.5\rangle$

Then:

$P(C{=}A|\text{HHHHTHHHHH}){=}\alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5{\approx}.00202\alpha$
$P(C{=}B|\text{HHHHTHHHHH}){=}\alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5{\approx}.00049\alpha$
$\mathbf{P}(C|\text{HHHHTHHHHH}){\approx}\langle .80, .20\rangle$

Probability of the Hidden Variable

$P(C{=}A|\mathrm{HHHHTHHHHH})$
$= \alpha P(\mathrm{HHHHTHHHHH}|C{=}A)P(C{=}A)$
$= \alpha P(\mathrm{H}|C{=}A)P(\mathrm{H}|C{=}A)\ldots P(\mathrm{H}|C{=}A)P(C{=}A)$
$= \alpha P(\mathrm{H}|A)P(\mathrm{H}|A)\ldots P(\mathrm{H}|A)P(C{=}A)$

Say we make an initial guess:

- $\mathbf{P}(C) = \langle 0.5, 0.5 \rangle$
- $\mathbf{P}(A) = \langle 0.6, 0.4 \rangle$
- $\mathbf{P}(B) = \langle 0.5, 0.5 \rangle$

Then:

$P(C{=}A|\mathrm{HHHHTHHHHH}) = \alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5 \approx .00202\alpha$
$P(C{=}B|\mathrm{HHHHTHHHHH}) = \alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5 \approx .00049\alpha$
$\mathbf{P}(C|\mathrm{HHHHTHHHHH}) \approx \langle .80, .20 \rangle$

Probability of the Hidden Variable

$$P(C{=}A|\mathrm{HHHHTHHHHH})$$
$$={\alpha}P(\mathrm{HHHHTHHHHH}|C{=}A)P(C{=}A)$$
$$={\alpha}P(\mathrm{H}|C{=}A)P(\mathrm{H}|C{=}A)\ldots P(\mathrm{H}|C{=}A)P(C{=}A)$$
$$={\alpha}P(\mathrm{H}|A)P(\mathrm{H}|A)\ldots P(\mathrm{H}|A)P(C{=}A)$$

Say we make an initial guess:

- $\mathbf{P}(C){=}\langle 0.5, 0.5\rangle$
- $\mathbf{P}(A){=}\langle 0.6, 0.4\rangle$
- $\mathbf{P}(B){=}\langle 0.5, 0.5\rangle$

Then:

$$P(C{=}A|\mathrm{HHHHTHHHHH}){=}\alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5{\approx}.00202\alpha$$
$$P(C{=}B|\mathrm{HHHHTHHHHH}){=}\alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5{\approx}.00049\alpha$$
$$\mathbf{P}(C|\mathrm{HHHHTHHHHH}){\approx}\langle .80, .20\rangle$$

Probability of the Hidden Variable

$$
\begin{aligned}
& P(C{=}A|\mathrm{HHHHTHHHHH}) \\
&= \alpha P(\mathrm{HHHHTHHHHH}|C{=}A)P(C{=}A) \\
&= \alpha P(\mathrm{H}|C{=}A)P(\mathrm{H}|C{=}A)\ldots P(\mathrm{H}|C{=}A)P(C{=}A) \\
&= \alpha P(\mathrm{H}|A)P(\mathrm{H}|A)\ldots P(\mathrm{H}|A)P(C{=}A)
\end{aligned}
$$

Say we make an initial guess:

- $\mathbf{P}(C) = \langle 0.5, 0.5 \rangle$
- $\mathbf{P}(A) = \langle 0.6, 0.4 \rangle$
- $\mathbf{P}(B) = \langle 0.5, 0.5 \rangle$

Then:

$$P(C{=}A|\mathrm{HHHHTHHHHH}) = \alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5 \approx .00202\alpha$$
$$P(C{=}B|\mathrm{HHHHTHHHHH}) = \alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5 \approx .00049\alpha$$
$$\mathbf{P}(C|\mathrm{HHHHTHHHHH}) \approx \langle .80, .20 \rangle$$

Probability of the Hidden Variable

$$
\begin{aligned}
&P(C{=}A|\text{HHHHTHHHHH})\\
&=\alpha P(\text{HHHHTHHHHH}|C{=}A)P(C{=}A)\\
&=\alpha P(\text{H}|C{=}A)P(\text{H}|C{=}A)\ldots P(\text{H}|C{=}A)P(C{=}A)\\
&=\alpha P(\text{H}|A)P(\text{H}|A)\ldots P(\text{H}|A)P(C{=}A)
\end{aligned}
$$

Say we make an initial guess:

- $\mathbf{P}(C){=}\langle 0.5, 0.5\rangle$
- $\mathbf{P}(A){=}\langle 0.6, 0.4\rangle$
- $\mathbf{P}(B){=}\langle 0.5, 0.5\rangle$

Then:

$P(C{=}A|\text{HHHHTHHHHH}){=}\alpha \cdot 0.6^9 \cdot 0.4^1 \cdot 0.5{\approx}.00202\alpha$

$P(C{=}B|\text{HHHHTHHHHH}){=}\alpha \cdot 0.5^9 \cdot 0.5^1 \cdot 0.5{\approx}.00049\alpha$

$\mathbf{P}(C|\text{HHHHTHHHHH}){\approx}\langle .80, .20\rangle$

Expectation Maximization Algorithm

1. Make initial guess of all probabilities $\mathbf{P}(X_0), \mathbf{P}(X_1), \ldots$
2. Calculate $\mathbf{P}(X_h|\text{data})$ for all hidden variables X_h
3. Tabulate partial counts for $X_0, X_1, \ldots$ from data
4. Normalize partial counts to get $\mathbf{P}(X_0), \mathbf{P}(X_1), \ldots$
5. Goto 2

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) = \alpha \cdot$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) = \alpha \cdot$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) = \alpha \cdot$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) = \alpha \cdot$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{matrix} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{matrix} \;\middle|\; bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$
- $P\left(\begin{matrix} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{matrix} \;\middle|\; bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{matrix} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{matrix} \;\middle|\; bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$
- $P\left(\begin{matrix} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{matrix} \;\middle|\; bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $b \Leftrightarrow y$:

- $P\left(\begin{matrix} \mathrm{b} \\ | \\ \mathrm{y} \end{matrix} \;\middle|\; b \Leftrightarrow y \right) = \alpha \cdot$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $b \Leftrightarrow y$:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y \right) = \alpha \cdot \frac{1}{2}$

EM in Machine Translation

Given parallel sentences

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

Make initial guess of all probabilities

- $P(b{=}x){=}\frac{1}{2}$, $P(b{=}y){=}\frac{1}{2}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $bc \Leftrightarrow xy$:

- $P\left(\begin{matrix}\mathrm{b} & \mathrm{c}\\ | & | \\ \mathrm{x} & \mathrm{y}\end{matrix} \;\middle|\; bc \Leftrightarrow xy\right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$
- $P\left(\begin{matrix}\mathrm{b} & & \mathrm{c}\\ & \times & \\ \mathrm{x} & & \mathrm{y}\end{matrix} \;\middle|\; bc \Leftrightarrow xy\right) = \alpha \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\alpha \Rightarrow \frac{1}{2}$

Calculate $\mathbf{P}(A|\text{data})$ for alignments A in $b \Leftrightarrow y$:

- $P\left(\begin{matrix}\mathrm{b}\\ | \\ \mathrm{y}\end{matrix} \;\middle|\; b \Leftrightarrow y\right) = \alpha \cdot \frac{1}{2} \Rightarrow 1$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{matrix} \text{b} & \text{c} \\ | & | \\ \text{x} & \text{y} \end{matrix} \;\middle|\; bc \Leftrightarrow xy \right) \Rightarrow \begin{matrix} \#(b = x) & \mathrel{+}= & \frac{1}{2} \\ \#(c = y) & \mathrel{+}= & \frac{1}{2} \end{matrix}$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{smallmatrix} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{smallmatrix} \;\middle|\; bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lll} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{smallmatrix} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{smallmatrix} \;\middle|\; bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lll} \#(b = y) & += & \frac{1}{2} \\ \#(c = x) & += & \frac{1}{2} \end{array}$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{ll} \#(b = x) & \mathrel{+}= \frac{1}{2} \\ \#(c = y) & \mathrel{+}= \frac{1}{2} \end{array}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{ll} \#(b = y) & \mathrel{+}= \frac{1}{2} \\ \#(c = x) & \mathrel{+}= \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y\right) \qquad \Rightarrow$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{rcl} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{c} \mathrm{b} \quad \mathrm{c} \\ \times \\ \mathrm{x} \quad \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{rcl} \#(b = y) & += & \frac{1}{2} \\ \#(c = x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y \right) \Rightarrow \#(b = y) \;\; += \;\; 1$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{smallmatrix} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{smallmatrix} \,\middle|\, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lll} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{smallmatrix} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{smallmatrix} \,\middle|\, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lll} \#(b = y) & += & \frac{1}{2} \\ \#(c = x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{smallmatrix} \mathrm{b} \\ | \\ \mathrm{y} \end{smallmatrix} \,\middle|\, b \Leftrightarrow y\right) \Rightarrow \#(b = y) \;+=\; 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x){=}$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ & \!\!\times\!\! \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b = y) & += & \frac{1}{2} \\ \#(c = x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y \right) \Rightarrow \#(b = y) \mathrel{+}= 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{3}{2}} = \frac{1}{4}$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{smallmatrix} \text{b} & \text{c} \\ | & | \\ \text{x} & \text{y} \end{smallmatrix} \,\middle|\, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{smallmatrix} \text{b} & & \text{c} \\ & \times & \\ \text{x} & & \text{y} \end{smallmatrix} \,\middle|\, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b = y) & += & \frac{1}{2} \\ \#(c = x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{smallmatrix} \text{b} \\ | \\ \text{y} \end{smallmatrix} \,\middle|\, b \Leftrightarrow y \right) \Rightarrow \#(b = y) \;\; += \;\; 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{1}{4}$, $P(b{=}y) =$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lcl} \#(b = x) & += & \frac{1}{2} \\ \#(c = y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lcl} \#(b = y) & += & \frac{1}{2} \\ \#(c = x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y\right) \Rightarrow \#(b = y) \;+=\; 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{1}{4}$, $P(b{=}y) = \frac{\frac{3}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{3}{4}$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lcl} \#(b=x) & += & \frac{1}{2} \\ \#(c=y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{c} \mathrm{b}\quad\mathrm{c} \\ \times \\ \mathrm{x}\quad\mathrm{y} \end{array} \middle| bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lcl} \#(b=y) & += & \frac{1}{2} \\ \#(c=x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| b \Leftrightarrow y\right) \Rightarrow \#(b=y) \mathrel{+}= 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{1}{4}$, $P(b{=}y) = \frac{\frac{3}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{3}{4}$
- $\mathbf{P}(c)$: $P(c{=}x) =$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lll} \#(b=x) & += & \frac{1}{2} \\ \#(c=y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lll} \#(b=y) & += & \frac{1}{2} \\ \#(c=x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y \right) \Rightarrow \#(b=y) \; += \; 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{1}{4}$, $P(b{=}y) = \frac{\frac{3}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{3}{4}$
- $\mathbf{P}(c)$: $P(c{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} = \frac{1}{2}$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \text{b} & \text{c} \\ | & | \\ \text{x} & \text{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lcl} \#(b=x) & += & \frac{1}{2} \\ \#(c=y) & += & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{ccc} \text{b} & & \text{c} \\ & \times & \\ \text{x} & & \text{y} \end{array} \middle| \, bc \Leftrightarrow xy\right) \Rightarrow \begin{array}{lcl} \#(b=y) & += & \frac{1}{2} \\ \#(c=x) & += & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \text{b} \\ | \\ \text{y} \end{array} \middle| \, b \Leftrightarrow y\right) \Rightarrow \#(b=y) \ += \ 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{1}{4}$, $P(b{=}y) = \frac{\frac{3}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{3}{4}$
- $\mathbf{P}(c)$: $P(c{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} = \frac{1}{2}$, $P(c{=}y) =$

EM in Machine Translation

Tabulate partial counts from $bc \Leftrightarrow xy$ alignments:

- $P\left(\begin{array}{cc} \mathrm{b} & \mathrm{c} \\ | & | \\ \mathrm{x} & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b=x) & \mathrel{+}= & \frac{1}{2} \\ \#(c=y) & \mathrel{+}= & \frac{1}{2} \end{array}$
- $P\left(\begin{array}{ccc} \mathrm{b} & & \mathrm{c} \\ & \times & \\ \mathrm{x} & & \mathrm{y} \end{array} \middle| \, bc \Leftrightarrow xy \right) \Rightarrow \begin{array}{lcl} \#(b=y) & \mathrel{+}= & \frac{1}{2} \\ \#(c=x) & \mathrel{+}= & \frac{1}{2} \end{array}$

Tabulate partial counts from $b \Leftrightarrow y$ alignments:

- $P\left(\begin{array}{c} \mathrm{b} \\ | \\ \mathrm{y} \end{array} \middle| \, b \Leftrightarrow y \right) \Rightarrow \#(b=y) \mathrel{+}= 1$

Normalize partial counts to get new probabilities:

- $\mathbf{P}(b)$: $P(b{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{1}{4}$, $P(b{=}y) = \frac{\frac{3}{2}}{\frac{1}{2}+\frac{3}{2}} = \frac{3}{4}$
- $\mathbf{P}(c)$: $P(c{=}x) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} = \frac{1}{2}$, $P(c{=}y) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} = \frac{1}{2}$

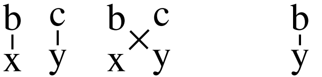
EM in Machine Translation Exercise

Given sentences:

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

With possible alignments:

$bc \Leftrightarrow xy$ $\qquad$ $b \Leftrightarrow y$



And current probabilities:

- $P(b{=}x){=}\frac{1}{4}$, $P(b{=}y){=}\frac{3}{4}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate the probabilities after the next iteration of EM

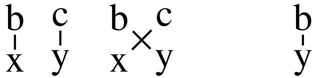
EM in Machine Translation Exercise

Given sentences:

- $bc \Leftrightarrow xy$
- $b \Leftrightarrow y$

With possible alignments:

$bc \Leftrightarrow xy$ $\qquad$ $b \Leftrightarrow y$



And current probabilities:

- $P(b{=}x){=}\frac{1}{4}$, $P(b{=}y){=}\frac{3}{4}$, $P(c{=}x){=}\frac{1}{2}$, $P(c{=}y){=}\frac{1}{2}$

Calculate the probabilities after the next iteration of EM

- $P(b{=}x){=}\frac{1}{8}$, $P(b{=}y){=}\frac{7}{8}$, $P(c{=}x){=}\frac{3}{4}$, $P(c{=}y){=}\frac{1}{4}$

Expectation Maximization Properties

Theoretical:

- Each iteration increases the log likelihood of the data
- Often guaranteed to converge to a local maximum
- Hill climbing, but no step size needed

Practical:

- Initialization can be critical
- May overfit to noise in the data

Expectation Maximization Properties

Theoretical:

- Each iteration increases the log likelihood of the data
- Often guaranteed to converge to a local maximum
- Hill climbing, but no step size needed

Practical:

- Initialization can be critical
- May overfit to noise in the data

Key Ideas

Naive Bayes

- Assume features conditionally independent given class
- Maximum likelihood estimates (i.e. count and divide)

Expectation Maximization

- Guess probabilities for all variables in model
- Calculate probabilities of hidden variables given data
- Tabulate partial counts $\Rightarrow$ new variable probabilities
- Iterate until local maximum is reached