Handwriting recognition

Character recognition, e.g., kernel SVMs
Webpage classification

Company home page vs Personal home page vs University home page vs …

Handwriting recognition 2

"c" more likely to come after an "a" than an "A"
Today – Bayesian networks

- One of the most exciting advancements in statistical AI in the last decades
- Generalizes naïve Bayes and logistic regression classifiers
- Compact representation for exponentially-large probability distributions
- Exploit conditional independencies
Causal structure

- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches
- How are these connected?

Possible queries

- Inference
  \[ P(F = 1 \mid N = 1) \]
- Most probable explanation
  \[ \max_{f, s, h} P(f, s, a, h \mid N = 1) \]
- Active data collection
  What variable should I observe next? 
  \( H = ? \), \( S = ? \)
Car starts BN

- 18 binary attributes
- Inference
  - P(BatteryAge|Starts=f)
  - \( P(Ba|s=c) = \sum_{r,e,f} P(q,r,f,e,...,r,e,f,s=c) \)
- 2^{16} terms, why so fast?
- Not impressed?
  - HailFinder BN – more than 3^{54} = 58149737003040059690390169 terms

Factored joint distribution - Preview

- \( P(F) \)
- P(A)
- P(S|F,A)
- P(H|S)
- P(N|S)

\( P(F,A,S,H,N) = P(F) P(A) P(S|F,A) P(H|S) P(N|S) \)

\( 2^{5} = 32 \) terms

9 parameters
What about probabilities?
Conditional probability tables (CPTs)

Flu
Allergy
Sinus
Headache
Nose

P(F) = 0.05
P(F|A) = 0.95
P(A) = 0.8

P(S|F, A) = 0.1
P(S|F, ¬A) = 0.3
P(S|¬F, A) = 0.2
P(S|¬F, ¬A) = 0.6

Number of parameters

Flu: 1 parameter
Allergy: 1 parameter
Sinus: 4 parameters (2 F, 2 A)
Headache: 2 parameters
Nose: 1 parameter

Total: 10 parameters

- more bias
- less flexible
- need less data to learn
- more accurate on similar domains
Key: Independence assumptions

Knowing sinus separates the variables from each other

(Marginal) Independence

- Flu and Allergy are (marginally) independent

\[
P(A) = P(A|F) = P(A|\neg F)
\]

\[
P(\neg A) = 1 - P(A)
\]

\[
P(\neg A|F) = P(\neg A|\neg F)
\]

\[
P(F,A) = P(F|A)P(A) = P(F|\neg A)P(\neg A)
\]

\[
P(\neg F,A) = P(\neg F|A)P(A) = P(\neg F|\neg A)P(\neg A)
\]
Marginally independent random variables

- **Sets** of variables $X$, $Y$
- $X$ is independent of $Y$ if
  - $P(F(X=x \perp Y=y), \forall x \in \text{Val}(X), y \in \text{Val}(Y))$
  - Shorthand: $\text{Marginal independence: } P(F(X \perp Y))$

**Proposition:** $P$ satisfies $(X \perp Y)$ if and only if
- $P(X,Y) = P(X) \cdot P(Y)$
  - $P(X|Y) = P(X)$ equivalent

Conditional independence

- Flu and Headache are not (marginally) independent
  - $P(H=\epsilon|F=\epsilon) \neq P(H=\epsilon)$
- Flu and Headache are independent given Sinus infection
  - $P(H=\epsilon|S=\epsilon) = P(H=\epsilon|S=\epsilon, F=\epsilon)$

More Generally:
- $X \perp Y|Z$
  - $P(X,Y|Z) = P(X|Z) \cdot P(Y|Z)$
  - $P(X|Y,Z) = P(X|Z)$ equivalent
Conditionally independent random variables

- **Sets** of variables \( X, Y, Z \)
- \( X \) is independent of \( Y \) given \( Z \) if
  \[
  P(x=x \mid y=y, z=z), \quad \forall x \in \text{Val}(X), \ y \in \text{Val}(Y), \ z \in \text{Val}(Z)
  \]
- Shorthand:
  - Conditional independence: \( P \vdash (X \perp Y \mid Z) \)
  - For \( P \vdash (X \perp Y \mid \emptyset) \), write \( P \vdash (X \perp Y) \)

**Proposition:** \( P \) satisfies \( (X \perp Y \mid Z) \) if and only if

\[
P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z)
\]
Explaining away

Local Markov Assumption:
A variable $X$ is independent of its non-descendants given its parents

Naïve Bayes revisited

Local Markov Assumption:
A variable $X$ is independent of its non-descendants given its parents