CS221 Problem Workout

Week 8
Outline

● **Bayesian networks: Learning**
  ○ Maximum likelihood
  ○ Smoothing
  ○ EM Algorithm

● **Problem discussion**
Bayesian networks: Learning

Given local probability distributions, i.e. $P(x | \text{parents}(x))$
Find conditional $P(Q | E=e)$

Given observations / samples
Find the local distributions, i.e. $P(x | \text{Parents}(x))$
Example

**Variables:**

- Genre $G \in \{\text{drama, comedy}\}$
- Rating $R \in \{1, 2, 3, 4, 5\}$

\[
\Pr(G = g, R = r) = p_G(g)p_R(r | g)
\]

\[\mathcal{D}_{\text{train}} = \{(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)\}\]

**Parameters:** $\theta = (p_G, p_R)$

Example borrowed from lecture slides
Outline

- Bayesian networks: Learning
  - Maximum likelihood
  - Smoothing
  - EM Algorithm
- Problem discussion
Maximum likelihood

**Input:** training examples $\mathcal{D}_{\text{train}}$ of full assignments

**Output:** parameters $\theta = \{ p_d : d \in D \}$

---

**Algorithm: count and normalize**

**Count:**
For each $x \in \mathcal{D}_{\text{train}}$:
  For each variable $x_i$:
    Increment $\text{count}_{d_i}(x_{\text{Parents}(i)}, x_i)$

**Normalize:**
For each $d$ and local assignment $x_{\text{Parents}(i)}$:
  Set $p_d(x_i \mid x_{\text{Parents}(i)}) \propto \text{count}_d(x_{\text{Parents}(i)}, x_i)$

---

Slide borrowed from lecture slides

Stanford University
Outline

- Bayesian networks: Learning
  - Maximum likelihood
  - Smoothing
  - EM Algorithm
- Problem discussion
Smoothing

Why?
- What if count is 0? Should P be 0?

How to solve?
- Initialize all counts with a non-zero constant $\lambda$

Observations
- Larger $\lambda$ -> more uniform distributions, less influenced by data
- Smaller $\lambda$ -> more influenced by data
- Infinite data -> effect of $\lambda$ vanishes
Final algorithm

Input: training examples $D_{train}$ of full assignments

Output: parameters $\theta = \{ p_d : d \in D \}$

Algorithm: count and normalize

Count:
For each $x \in D_{train}$:
For each variable $x_i$:
Increment $\text{count}_{d_i}(x_{\text{Parents}(i)}, x_i)$

Normalize:
For each $d$ and local assignment $x_{\text{Parents}(i)}$:
Set $p_d(x_i | x_{\text{Parents}(i)}) \propto \text{count}_d(x_{\text{Parents}(i)}, x_i) + \lambda$
Outline

- Bayesian networks: Learning
  - Maximum likelihood
  - Smoothing
  - EM Algorithm
- Problem discussion
EM Algorithm

Variables: $H$ is hidden, $E = e$ is observed

Example:

Maximum marginal likelihood objective:

$$\max_\theta \prod_{e \in \mathcal{D}_{train}} \mathbb{P}(E = e; \theta)$$

$$= \max_\theta \prod_{e \in \mathcal{D}_{train}} \sum_h \mathbb{P}(H = h, E = e; \theta)$$

Slide borrowed from lecture slides

Stanford University
EM Algorithm

Initialize $\theta$ randomly

Until convergence:

# E Step
for each $e$ in Data:
  for each $h$:
    $q(h; e) = P(H = h \mid E = e; \theta)$ … inference
# Update table from $\{e, \text{count}(e)\}$ to $\{(h,e), (q(h; e) \times \text{count}(e))\}$
# Now no variables are hidden

# M step
update($\theta$) using Table $\{(h,e), (q(h; e) \times \text{count}(e))\}$ … MLE
Summary

- Given data learn the parameters of bayesian net

**MLE**
\[ p \propto \text{count}(x_i; \text{parents}(x_i)) \]

**Smoothing**
\[ p \propto \text{count}(x_i; \text{parents}(x_i)) + \lambda \]

**EM**
Data is *incomplete*

*E step: compute counts*

*M step: MLE*
Outline

● Bayesian networks: Learning
  ○ Maximum likelihood
  ○ Smoothing
  ○ EM Algorithm

● Problem discussion
Problem: P2, Winter 2021 Exam 2

Gambling Machine

Pick a coin Toss it

P_0(C_x) = \lambda_0

With probability \lambda coin at “t” is same as “t-1”

P_C(H) = p_C
How does the bayesian net look?

\[ P(c_i) = \begin{cases} \theta_0 & \text{if } c_i = x \\ 1-\theta_0 & \text{else} \end{cases} \]

\[ P(c_{i+1} | c_i) = \begin{cases} \theta & \text{if } c_{i+1} = c_i \\ 1-\theta & \text{else} \end{cases} \]

\[ P(o_i | c_i) = \begin{cases} P_{c_i} & \text{if } o_i = H \\ 1-P_{c_i} & \text{else} \end{cases} \]
Learning using EM algorithm

- Data = \{H, H, T\}
- \(\lambda_0\) and \(\lambda\) are given. To find: \(p_X\) and \(p_Y\)

Why do we need EM?
- What is not observed?
- \(C_i\) is not observed

How do we use EM?
- Compute \(q(c_i)\) using \(p'_X\) and \(p'_Y\)
- Use ML to update \(p'_X\) and \(p'_Y\)
Given q’s compute update

<table>
<thead>
<tr>
<th></th>
<th>T=1</th>
<th>T=2</th>
<th>T=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0.1</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Y</td>
<td>0.9</td>
<td>0.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Data = {H, H, T}

Compute:

<table>
<thead>
<tr>
<th>$C_i$</th>
<th>$O_i$</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Given q’s compute update

<table>
<thead>
<tr>
<th></th>
<th>T=1</th>
<th>T=2</th>
<th>T=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0.1</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Y</td>
<td>0.9</td>
<td>0.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Data = \{H, H, T\}

<table>
<thead>
<tr>
<th>C_i</th>
<th>O_i</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>H</td>
<td>0.1</td>
</tr>
<tr>
<td>X</td>
<td>H</td>
<td>0.5</td>
</tr>
<tr>
<td>X</td>
<td>T</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[
\text{P}(H \mid X) = p'_{x} = \frac{(0.1 + 0.5)}{(0.1 + 0.5 + 0.3)} \quad \ldots \text{MLE}
\]
Thank You