**Section A: HMMs + BayesNets**

**Viterbi Algorithm (edge weights)**

- \( e_t \) \( \rightarrow \) \( T \rightarrow T+1 \)
- \( \Delta \)
- \( \Delta \)
- \( \Delta \)

For \( t = 1, ..., T+1 \):
- For \( k = 1, ..., K \):
  - \( W_t(k) = \max_{j \in \{1, ..., K\}} \{ W_{t-1}(j) s_{jkt} \} \)
  - \( b_t(k) = \text{argmax}_{j \in \{1, ..., K\}} \{ W_{t-1}(j) s_{jkt} \} \)

**Example: Senate Race in PA**

- \( F = 1 \Rightarrow \) Fetterman beats O2
- \( B = 1 \Rightarrow \) politics is boring
- \( C = 1 \Rightarrow \) voters are finally Gavin
- \( L = 1 \Rightarrow \) voters are finally Liberal
- \( S = 1 \Rightarrow \) politics is Sensationalized
- \( D = 1 \Rightarrow \) voters are concerned about Fetterman's disability
- \( E = 1 \Rightarrow \) voters are concerned about O2's explanation of how w/ disabilities
- \( M = 1 \Rightarrow \) millennials are involved

**Question:** How to represent \( P(F, B, C, L, S, D, E) \)?

**Idea #1: Chain Rule**

\[
\]

Con: not very compact

**Idea #2: Full Joint Table**

<table>
<thead>
<tr>
<th>F</th>
<th>B</th>
<th>C</th>
<th>L</th>
<th>S</th>
<th>D</th>
<th>E</th>
<th>( p(*) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \Theta_1 )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>( \Theta_2 )</td>
</tr>
</tbody>
</table>

Con: not very compact
**Idea #3: Complete Independence**

\[ P(F, B, \ldots, E) > P(F)P(B) \ldots P(E) \]
- **Pros:** compact
- **Cons:** not very expensive

**Idea #4: Naive Bayes**

\[ P(F, B, \ldots, E) = P(B|F) \ldots P(E|F)P(F) \]

**Idea #5: Bayesian Network**

---

**Proof of Cond. Indep.**

**Case #1: Cascade**

\[ p(x, z|y) = \frac{p(x, y, z)}{p(y)} \]
\[ = \frac{p(x|y)p(y|z)P(z)}{p(y)} \]
\[ = \frac{p(x|y)p(z|y)p(y)}{p(y)} \]
\[ = p(x|y)p(z|y) \]
\[ \Rightarrow x \perp z \mid y \]

**Case #2: Common Parent** \( \Rightarrow \) left as exercise

**Case #3: V-structure**

\[ w_m - w'_m = f(\hat{y}(i) - P(y=1|x(i))) \cdot x(i) \]
\[
\omega_m = \omega_m - \delta \left( y^{(i)} - \frac{\mathcal{P}(y=1|x^{(i)})}{\mathcal{P}(y=0|x^{(i)})} x^{(i)} \right)
\]

\[
\sigma(\omega^T x + b)
\]

**Likelihood:**

\[
p(D | \theta) \propto \prod_{i=1}^{N} p(x^{(i)} | \theta) \prod_{i=1}^{N} p(y^{(i)} | x^{(i)}, \theta)
\]

\[
L_{y,x}(\theta)
\]

**Conditional Likelihood:**

\[
p(y_2 = n | x_1, x_2, x_3) = \sum_{y_1} \sum_{y_3} p(y_1, y_2 = n | x_1, x_2, x_3)
\]

\[
= \sum_{y_1} \sum_{y_3} p(y_1, y_2 = n | y_3 = y_3 | x_1, x_2, x_3)
\]

\[
= \sum_{y_1} \sum_{y_3} p(y_1, y_2 = n | y_3 = y_3 | x_1, x_2, x_3)
\]
\[ p(y_2 | x_1, x_2, x_3) = \sum_{y_1} \sum_{y_3} p(y_1, y_2, y_3 | x_1, x_2, x_3) \]

\[ = \sum_{y_1} \sum_{y_2} \sum_{x_4} \sum_{x_5} p(y_1, y_2, y_3, x_4, x_5 | x_1, x_2, x_3) \]

\[ \neq p(y_2 | x_1, x_2, x_3, x_4, x_5) \]

\[ A \perp B | C \]

\[ p(A, B, C, D, E) = p(A | C) p(B | C) p(C | D, E) p(D) p(E) \]

\[ \frac{\partial h}{\partial w} \in \mathbb{M} \]

\[ w \in \mathbb{R}^{M \times D} \]

\[ \frac{\partial h_i}{\partial x_j} = \mathbb{N}(i, j) \]

\[ x \in \mathbb{R}^N \]

\[ \tilde{g} = s(t \cdot (x)) = o(x) \]

\[ \frac{\partial h_i}{\partial x_i} = h_i (1 - h_i) \]