# Hidden Markov Models I

Notation

- r.v.'s: $X, Y, Z$
- Events: $A, B, C$
- Probability: $P\{A\}$, $P\{A, B\}$ or $P\{A \cap B\}$, $P\{A \cup B\}$
- Conditional probability: $P\{A|B\}$ ("A given B")

$$P\{A|B\} = \frac{P\{A, B\}}{P\{B\}}$$

  Equivalently,

$$P\{A, B\} = P\{A|B\} P\{B\}$$

- $P\{X = x\}$, $P\{X = x | Y = y\}$ where $x, y$ are fixed.

# Hidden Markov Models II

- Markov chain: A set $\{1, \ldots, N\}$ is called the *state space* and denoted by $S$. A collection of r.v.'s $\{X_t\}_{t=0}^{\infty}$ with transition probability matrix $P = (p_{ij})_{N \times N}$. $p_{ij} = P\{X_t = j | X_{t-1} = i\}$. Initial state probabilities $\pi_i = P\{X_1 = i\}$ for all states $i$.

Hidden Markov Models (HMMs)
Illustrated on the class example.

- Hidden: states 1, 2, 3. State variables $\{X_t\}_{t=0}^{\infty}$ (Markov chain). Transition probability matrix $P = (p_{ij})$.
- Observable: Observations $\{R, G, B\}$ (observation space). Observation variables $\{Y_t\}_{t=0}^{\infty}$. Emission probabilities

$$Q = (q_{jk}) = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}. \quad q_{jk} = P\{Y_t = k | X_t = j\}.$$

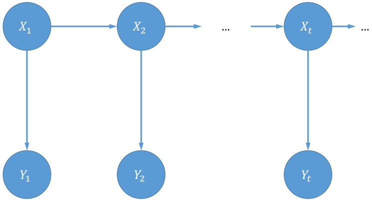


Figure 1: Bayesian Network Diagram

- Bayesian network diagram (dependency between $X_t$'s and $Y_t$'s: See Figure 1.

Three Questions to be answered:

- **Q1:** Given $P, Q$, find $P\{Y_1 = y_1, \ldots, Y_T = y_T\}$

- **Q2:** Given $P$, $Q$, find

$$
\begin{aligned}
&(x_1^*, \ldots, x_T^*) \\
&= \underset{(x_1,\ldots,x_T)}{\arg\max}\, P\{X_1 = x_1, \ldots, X_T = x_T | Y_1 = y_1, \ldots, Y_T = y_T\}
\end{aligned}
$$

- **Q3:** How to estimate $P$, $Q$ from data? A further question would be how to determine the number of states ($N$) and the number of observations ($M$).

Algorithms:

**Q1:** A naive method. By Total Probability Theorem,

$$
\begin{aligned}
&P\{Y_1 = y_1, \ldots, Y_T = y_T\} \\
&= \sum_{(x_1,\ldots,x_T)\in S^T} P\{Y_1 = y_1, \ldots, Y_T = y_T | X_1 = x_1, \ldots, X_T = x_T\} \\
&\qquad \cdot P\{X_1 = x_1, \ldots, X_T = x_T\}
\end{aligned}
$$

reasonone.ai

# Hidden Markov Models V

This provides a solution to **Q1**, as

$$P\{Y_1 = y_1, \dots, Y_T = y_T | X_1 = x_1, \dots, X_T = x_T\} = q_{x_1,y_1} \cdots q_{x_T,y_T}$$

$$P\{X_1 = x_1, \dots, X_T = x_T\} = \pi_{x_1} p_{x_1,x_2} p_{x_2,x_3} \cdots p_{x_{T-1},x_T},$$

where the preceding equation follows from an iterated application of condition probability and the Markov property. There are $N^T$ different state sequences $(x_1, \dots, x_T)$. For our example, the computational complexity is $3^T$, which grows exponentially in $T$ (definitely not desirable!).

A improved method based on recursion.
By Total Probability Theorem, we get

$$P\{Y_1 = y_1, \ldots, Y_T = y_T\} = \sum_{x_T \in S} P\{Y_1 = y_1, \ldots, Y_T = y_T, X_T = x_T\}$$

Define $f_t(x_t) = P\{Y_1 = y_1, \ldots, Y_t = y_t, X_t = x_t\}$ for all $t$.

$$f_T(x_T) = P\{Y_T = y_T | Y_1 = y_1, \ldots, Y_{T-1} = y_{T-1}, X_T = x_T\}$$
$$\cdot P\{Y_1 = y_1, \ldots, Y_{T-1} = y_{T-1}, X_T = x_T\}$$
$$= q_{x_T, y_T} P\{Y_1 = y_1, \ldots, Y_{T-1} = y_{T-1}, X_T = x_T\}$$
$$= q_{x_T, y_T} \sum_{x_{T-1}} P\{Y_1 = y_1, \ldots, Y_{T-1} = y_{T-1}, X_{T-1} = x_{T-1}, X_T = x_T\}$$
$$= q_{x_T, y_T} \sum_{x_{T-1} \in S} p_{x_{T-1}, x_T} f_{T-1}(x_{T-1})$$

Thus, we found a recursion.
The basis step:

$$f_1(x_1) = P\{Y_1 = y_1, X_1 = x_1\} = \pi_{x_1} q_{x_1, y_1}, \ \forall x_1 \in S$$

A forward algorithm with complexity $O(N^2 T)$

# Hidden Markov Models VIII

**Q2:** Find $(x_1^*, \ldots, x_T^*) = \arg\max_{(x_1,\ldots,x_T)} P\{X_1 = x_1, \ldots, X_T = x_T | Y_1 = y_1, \ldots, Y_T = y_T\}$
Consider the equivalent optimization problem

$$\max_{(x_1,\ldots,x_T)} P\{X_1 = x_1, \ldots, X_T = x_T, Y_1 = y_1, \ldots, Y_T = y_T\}$$

$$\Leftrightarrow$$

$$\max_{(x_1,\ldots,x_T)} (\pi_{x_1} p_{x_1,x_2} \ldots p_{x_{T-1},x_T})(q_{x_1,y_1} \ldots q_{x_T,y_T})$$

Use dynamic programming:
Define

$$f_T(x_T) = q_{x_T,y_T} \max_{(x_1,\ldots,x_{T-1})} (\pi_{x_1} p_{x_1,x_2} \ldots p_{x_{T-1},x_T})(q_{x_1,y_1} \ldots q_{x_{T-1},y_{T-1}})$$

# Hidden Markov Models IX

Then,

$$f_T(x_T) = q_{x_T,y_T} \max_{x_{T-1}\in S} p_{x_{T-1},x_T} f_{T-1}(x_{T-1})$$

Then original problem can be solved by solving

$$\max_{x_T\in S} f_T(x_T)$$

Recursion:

$$f_T(x_T) = q_{x_T,y_T} \max_{x_{T-1}\in S} p_{x_{T-1},x_T} f_{T-1}(x_{T-1})$$

The basis step:

$$f_1(x_1) = q_{x_1,y_1}, \ \forall x_1 \in S$$

For the detailed DP algorithm (due to Viterbi), cf. Example 3 Distributing Scientists to Research Teams (Hillier & Lieberman pp. 440–442)

# Hidden Markov Models X

**Q3:** How to estimate $P, Q$ from data?
In actuality, will need to estimate $N$ as well as $P, Q$; to simplify exposition, only consider the estimation of $P, Q$.

The general approach:

- split the data into two parts, one for training and the other for validation;
- make an initial guess of $P, Q$ based on the training data;
- fit the model $(P, Q)$ to the training data using maximum likelihood principle, i.e., to maximize the probability in **Q1**.

# Hidden Markov Models XI

Matlab demo:

http://www.mathworks.com/help/stats/hidden-markov-models-hmm.html

# Hidden Markov Models XII

Numerical Example:

$$P = \begin{pmatrix} .6 & .2 & .2 \\ .1 & .3 & .6 \\ .3 & .1 & .6 \end{pmatrix} \qquad Q = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 1/6 & 1/6 & 2/3 \end{pmatrix}$$