EE 126 — Probability and Random Processes

April 24, 2018 — Hidden Markov Models

1 Lecture

There are many forms of inference:

- **Filtering:** $Y_0, Y_1, ..., Y_T \to \boxed{\text{Filter}} \to \hat{X}_T$
  - e.g. tracking in real-time
- **Predicting:** $Y_0, Y_1, ..., Y_T \to \boxed{\text{Predict}} \to \hat{Y}_{T+1}$
  - Given data from 0 to T, we want to predict what's going to happen tomorrow.
  - e.g. radar tracking, stock prices
- **Smoothing:** $Y_0, Y_1, ..., Y_T \to \boxed{\text{Smooth}} \to \hat{X}_t \quad (t \le T)$
  - This is about being able to estimate in non-real time.
  - We get all this data and we chew on it, and maybe at time T we want a better estimate of what happened at time 1. We don't mind waiting; we just really want to know what was happening at time 1.
  - *Non-causal* applications: record data and then want to know "where was the murderer?"
  - e.g. inferring things in the past (non-causal filtering), like the cause of a car crash
  - The main difference between smoothing and filtering is *real-time* versus *non real-time*.
- **Max Likelihood State Estimation (MLSE):** $Y_0, Y_1, ..., Y_T \to \boxed{\text{MLSE}} \to \{\hat{X}_0, \hat{X}_1, ..., \hat{X}_T\}$
  - Here, we'd like to find the sequence of states that best explains what we observed.
  - e.g. convolutional coding (Viterbi algorithm), auto-spell

HMMs

An HMM is a random sequence $\{x_n, y_n\}$ where $x_n \in \mathcal{X} = \{1, 2, ..., N\}$ and $y_n \in \mathcal{Y} = \{1, 2, ..., M\}$, for which

$$P(X_0 = x_0, ..., X_n = x_n, Y_0 = y_0, ..., Y_n = y_n) = \pi_0(x_0) \cdot Q(x_0, y_0) \cdot P(x_0, x_1) \cdot Q(x_1, y_1) \cdots P(x_{n-1}, x_n) \cdot Q(x_n, y_n)$$

where

π_0 is the initial state distribution

Q is the [observation] emission probability

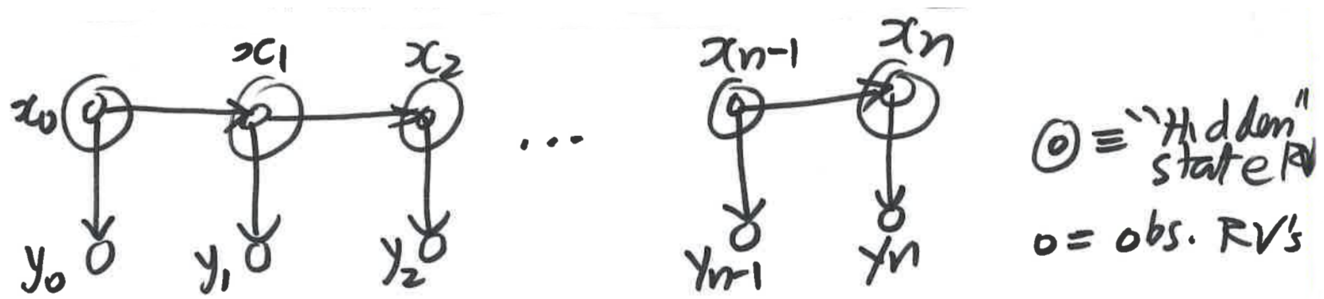
P is the transition probability

A diagram of the HMM setup can be seen on the next page.

We have hidden states x_i and observable states (emissions) y_i. The goal is to make a maximum likelihood sequence estimate of the hidden states x_i given the observations y_i.

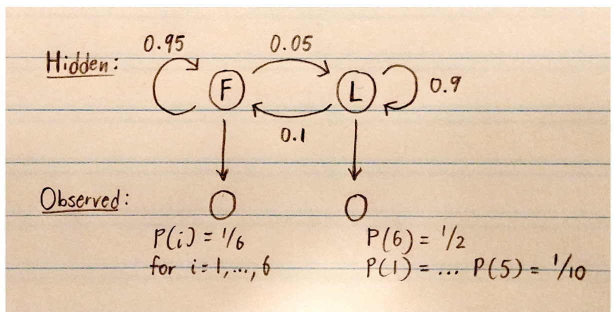
Example: $n = 2$.

$$P(x_0, y_0, x_1, y_1) = P(x_0)P(y_0 \mid x_0)P(x_1 \mid x_0)P(y_1 \mid x_1)$$



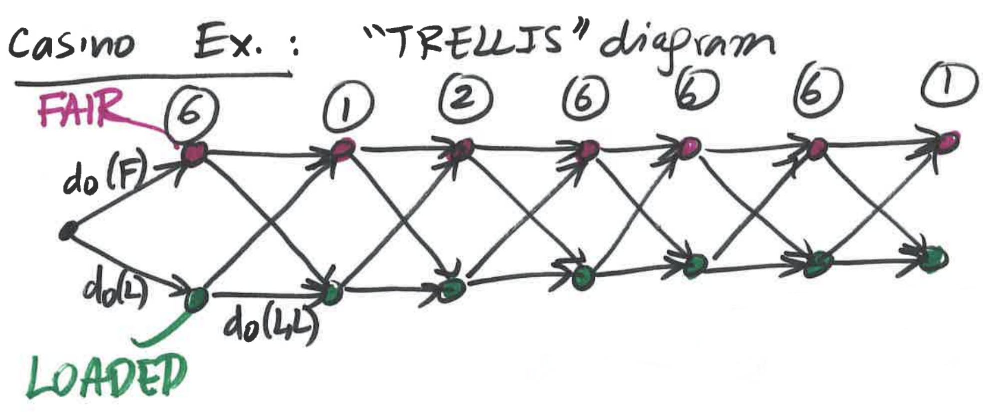
source: notes from Kannan Ramchandran's EE 126 lecture

Example: "nearly honest" casino. A casino uses a fair die most of the time, but switches to a loaded die as needed.



Problem: given the observations (a sequence of die rolls), find the most likely sequence of hidden states (die labels). In other words, find $\mathrm{MAP}[X^n \mid Y^n = y^n]$.

$$\begin{aligned}
x^{n^*} &= \arg\max_{x^n \in \mathcal{X}^n} P(X^n = x^n \mid Y^n = y^n) \\
&= \arg\max_{x^n \in \mathcal{X}^n} \pi_0(x_0)Q(x_0, y_0)P(x_0, x_1)\cdots P(x_{n-1}, x_n)Q(x_n, y_n) \\
&= \arg\max_{x^n \in \mathcal{X}^n} \underbrace{\log\left[\pi_0(x_0)Q(x_0, y_0)\right]}_{\text{define as } -d_0(x_0)} + \sum_{m=1}^{n} \underbrace{\log\left[P(x_{m-1}, x_m)Q(x_m, y_m)\right]}_{\text{define as } -d_m(x_{m-1}, x_m)} \\
&= \arg\min_{x^n \in \mathcal{X}^n} d_0(x_0) + \sum_{m=1}^{n} d_m(x_{m-1}, x_m)
\end{aligned}$$



source: notes from Kannan Ramchandran's EE 126 lecture

The numbers are the observations. Now, using probabilities as distances, we just need to find the *shortest path* from the start to the end state. Use Bellman-Ford. That's all there is to it.