Class Tutorial 4

1. Markov Chains

Definition: the hitting time T_{ij} is a random variable that corresponds to the first time of visiting state j when starting from state i. The probability function of T_{ij} is given by

$$P\left(T_{ij}=k\right)=P\left(X_k=j,X_{k-1}\neq j,\ldots,X_1\neq j\,\middle|\,X_0=i\right)\qquad k=1,2,\ldots$$

Also, let $\mu_{ij}=\mathbb{E}\left[T_{ij}\right]$.

Consider the two state Markov chain with transition matrix

$$\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

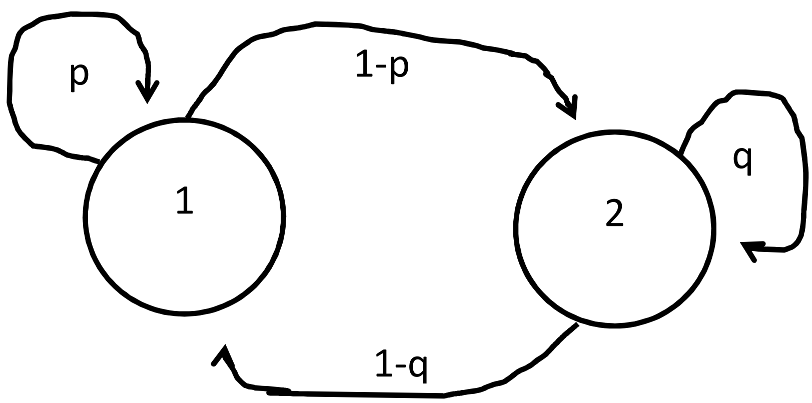
a. Draw the transition diagram for this Markov chain.

b. Calculate μ_{ij} directly, using the probability function.

c. Calculate μ_{ij} by conditioning, using the smoothing theorem.

Solution

a.



b. We will calculate μ_{12}. The calculation is similar for the other terms.

By definition:

$$P(T_{12}=k)=P(X_k=2,X_{k-1}=1,\dots,X_1=1|X_0=1)$$
$$=p^{k-1}(1-p)$$

And

$$\mu_{12}=\sum_{k=1}^{\infty}kp^{k-1}(1-p)=(1-p)\sum_{k=0}^{\infty}kp^{k-1}$$
$$=(1-p)\sum_{k=0}^{\infty}\frac{\partial}{\partial p}p^k=(1-p)\frac{\partial}{\partial p}\left(\frac{1}{1-p}\right)$$
$$=\frac{(1-p)}{(1-p)^2}=\frac{1}{1-p}$$

c. Note that by smoothing we have $\mathbb{E}[T_{12}]=\mathbb{E}\left[\mathbb{E}\left[T_{12}|X_1\right]\right]$,

but, observe that by definition we have that

$$\mathbb{E}\left[T_{12}|X_1=1\right]=\mathbb{E}[T_{12}]+1 \ ,$$

And

$$\mathbb{E}\left[T_{12}|X_1=2\right]=1,$$

Therefore

$$\mathbb{E}[T_{12}]=\mu_{12}=p(\mu_{12}+1)+(1-p)\cdot 1 \ .$$

Solving for μ_{12} gives

$$\mu_{12}=\frac{1}{1-p} \ .$$

2. Viterbi Algorithm

Consider the following Hidden Markov Model definition. A sequence of T states $X_1,\dots,X_T\in\mathcal{X}$ are generated by a Markov chain with transition probabilities $P(X_{k+1}=j\,|\,X_k=i)=p_{ij}$ and initial state distribution $P(X_1=i)=\rho_i$. For each state X_k, an observation $Y_k\in\mathcal{Y}$ is generated according to a distribution $P(Y=j\,|\,X=i)=q_{ij}$, and independently of other states or observations.

The HMM decoding problem is as follows. Given an *observation* sequence $Y_1,\dots,Y_T$, and the HMM parameters p_{ij},ρ_i , and q_{ij} , find the most likely state sequence $X_1,\dots,X_T$ that generated them, i.e., find

$$\max_{X_1,\dots,X_T} P\left(X_1,\dots,X_T \middle| Y_1,\dots,Y_T\right)$$

Using Bayes rule, we have

$$\begin{aligned}\max_{X_1,\dots,X_T} P\left(X_1,\dots,X_T \middle| Y_1,\dots,Y_T\right) &= \max_{X_1,\dots,X_T} \frac{P\left(X_1,\dots,X_T,Y_1,\dots,Y_T\right)}{P\left(Y_1,\dots,Y_T\right)} \\ &= \max_{X_1,\dots,X_T} P\left(X_1,\dots,X_T,Y_1,\dots,Y_T\right)\end{aligned}$$

The Viterbi algorithm uses dynamic programming to solve the HMM decoding problem.

a. Write $P\left(X_1,\dots,X_T,Y_1,\dots,Y_T\right)$ explicitly using p_{ij},ρ_i , and q_{ij} .

Let $V_t\left(i\right)$ denote the likelihood of the most likely sequence that generated $Y_1,\dots,Y_t$ and ends in state $X_t = i$, i.e.,

$$V_t(i) = \max_{X_1,\dots,X_{t-1}} P\left(X_1,\dots,X_{t-1},Y_1,\dots,Y_t,X_t = i\right)$$

b. Write an expression for $V_1\left(i\right)$.

c. Write a recursive formula for $V_{t+1}\left(i\right)$.

d. What is the complexity of the resulting recursive algorithm?

Solution:

a. $P\left(X_1,\dots,X_T,Y_1,\dots,Y_T\right) = \rho_{X_1} q_{X_1,Y_1} p_{X_1,X_2} \cdots p_{X_T,X_{T-1}} q_{X_T,Y_T}$

b.

$$V_1\left(i\right) = P\left(Y_1, X_1 = i\right) = \rho_i q_{i,Y_1}$$

c.

$$
\begin{aligned}
V_{t+1}(i) &= \max_{X_1,\ldots,X_t} P\left(X_1,\ldots,X_t,Y_1,\ldots,Y_{t+1},X_{t+1}=i\right) \\
&= \max_{X_1,\ldots,X_t} P\left(X_{t+1}=i,Y_{t+1} \mid X_t\right) P\left(X_1,\ldots,X_t,Y_1,\ldots,Y_t\right) \\
&= \max_{X_t} \max_{X_1,\ldots,X_{t-1}} P\left(X_{t+1}=i,Y_{t+1} \mid X_t\right) P\left(X_1,\ldots,X_t,Y_1,\ldots,Y_t\right) \\
&= \max_{X_t} P\left(X_{t+1}=i,Y_{t+1} \mid X_t\right) \max_{X_1,\ldots,X_{t-1}} P\left(X_1,\ldots,X_t,Y_1,\ldots,Y_t\right) \\
&= \max_{X_t} P\left(X_{t+1}=i,Y_{t+1} \mid X_t\right) V_t\left(X_t\right) \\
&= \max_{X_t} p_{X_t,i} q_{i,Y_{t+1}} V_t\left(X_t\right) \\
&= q_{i,Y_{t+1}} \max_{X_t} p_{X_t,i} V_t\left(X_t\right)
\end{aligned}
$$

d. The complexity is $\mathcal{O}\left(T\left|\mathcal{X}\right|^2\right)$.