Homework # 8

Reading

- Section 13.2 introduces hidden Markov models. The previous section, 13.1, discuss Markov chains and similar models of sequential random variables. In problem 1 below, we use $X(t)$ as the hidden state. Bishop instead uses Z for the hidden state, following the notation for EM, which we will also pick up next week.

- I've attached the original diffusion maps paper by Coifman and Lafon. The wikipedia article on diffusion maps follows the Coifman and Lafon paper closely and is worth reading,

  `https://en.wikipedia.org/wiki/Diffusion_map`

1. (This problem is based on the cheating-casino hidden Markov model we discussed in class.) Let $X(t)$ be a Markov chain on the state space $\{F, C\}$ (F - fair, C - cheating). Suppose that $X(t)$ changes state with probability $\alpha = .05$ regardless of its current state. Let $Y(t)$ be a r.v. with values from $\{1, 2, 3, 4, 5, 6\}$. ($Y(t)$ corresponds to the t-th role of a die). If $X(t) = F$ then $Y(t)$ is uniformly distributed on $\{1, 2, 3, 4, 5, 6\}$ (a fair die). If $X(t) = C$ then $P(Y(t) = 6) = 1/50$ and all values for $Y(t)$ are equally likely. Assume $X(0) = F$ where π is the stationary distribution of $X(t)$.

   (a) Write a function, **SampleCasino(T)** that samples $X(t)$, $Y(t)$ for $t \leq T$.

   (b) Use your simulation from part (a) to produce a single realization of $X(t)$ and $Y(t)$ up to time step $T = 200$. Pretend that you don't know the $X(t)$ values, but that you know the $Y(t)$ values generated. Let $(j_0, j_1, \ldots, j_T)$ be the sequence of $Y(t)$ values you generated. Given a sequence $(i_0, i_1, \ldots, i_T)$ of states $i \in \{F, C\}$, write an expression for the probability,

$$P(X(0) = i_0, X(1) = i_1, \ldots, X(T) = i_T \mid Y(0) = j_0, Y(1) = j_1, \ldots, Y(T) = j_T) \quad (1)$$

Set $\alpha = P(Y(0) = j_0, Y(1) = j_1, \ldots, Y(T) = j_T)$. Your expression for (1) should be a function of α and the i_s, j_s for $s = 0, 1, 2, \ldots, T$. Provide an expression for α (you can express α through $T+1$ sums).

(c) Let $\nu(i_0, i_1, \ldots, i_T)$ be the conditional probability given in (1). Let Z be the r.v. with distribution ν. What is the state space of Z? (We discussed this in class.)

(d) Using a Metropolis-Hastings approach, construct a Markov chain $W(s)$ that has Z as its stationary distribution. (Here I'll use s as the time variable so as not to confuse it with the t variable of $X(t)$). Use your sampler to estimate $P(X(t) = C \mid Y(0) = j_0, Y(1) = j_1, \ldots, Y(T) = j_T)$ where t is a given value. Using a single long run of $W(s)$, estimate $P(X(t) = C \mid Y(0) = j_0, Y(1) = j_1, \ldots, Y(T) = j_T)$, don't forget to include a burn-in time. Do this for all $t \leq 200$. You can use a single long run of $W(s)$ for each value of t. Plot $P(X(t) = C \mid Y(0) = j_0, Y(1) = j_1, \ldots, Y(T) = j_T)$ as a function of t and compare the probabilities you computed to the actual state of the casino.

2. Attached you will find a R script `make_1d_manifold.R` that constructs data points $x^{(i)} \in \mathbb{R}^{10}$ for $i = 1, 2, \ldots, 500$ that are localized around a 1-d manifold. The data points produced by the script are in `diffusion_maps_data.csv`.

(a) Look at the file `diffusion_maps_data.csv`. The data points are given in the first 10 columns. The 11th column gives a parameter β discussed in the next subproblem. Can you find a pattern in the data? (The answer will be no, I think.)

(b) Read the script and describe what the 1-d manifold looks like. Associated with each data point is a scalar $\beta \in [0, 1]$. Explain how β paramatrizes the manifold.

(c) Reduce the data to $\mathbb{R}^2$ and $\mathbb{R}$ using PCA. Plot the data points in the reduced dimension and use color to represent the value of β.

(d) i. The authors of the diffusion maps paper (see Reading above) introduce the diffusion distance between the data points $x^{(i)}$, $x^{(j)}$,

$$D_t^2(x^{(i)}, x^{(j)}) = \sum_{k=1}^{N} \Big[(P(Z(t) = x^{(k)} \mid Z(0) = x^{(i)}) - P(Z(t) = x^{(k)} \mid Z(0) = x^{(j)}\Big]^2 \frac{1}{\pi(x^{(k)})} \tag{2}$$

Explain the intuition behind this distance. How is $Z(t)$ constructed? What is π? Here I'm not looking for any proofs or derivations. Just explain your understanding of the construction.

ii. The authors describe a mapping from $x^{(i)} \in \mathbb{R}^n$ to $y^{(i)} \in \mathbb{R}^N$, where $y^{(i)}$ is given by

$$y^{(i)} = \begin{pmatrix} \lambda_1^t r_i^{(1)} \\ \lambda_2^t r_i^{(2)} \\ \vdots \\ \lambda_N^t r_i^{(N)} \end{pmatrix} \tag{3}$$

Explain how to compute the λ and r. What is the formula for $D_t^2(x^{(i)}, x^{(j)})$ in terms of $y^{(i)}, y^{(j)}$? Again, I'm not looking for proofs or step-by-step derivations. Just show my how you would do the computations and state the formula. (The lecture video and paper provide step by step derivations.)

iii. Describe how we would use diffusion maps to dimensionally reduce the data.

iv. Now repeat (c), but use diffusion maps to dimensionally reduce the data. Experiment with different kernels and different time parameters t in the diffusion maps.