CS 189/289A Introduction to Machine Learning
Fall 2021 Jennifer Listgarten and Jitendra Malik **DIS12**

1 Probabilistic Graphical Models

Recall that we can represent joint probability distributions with directed acyclic graphs (DAGs). Let G be a DAG with vertices $X_1, ..., X_k$. If P is a (joint) distribution for $X_1, ..., X_k$ with (joint) probability mass function p, we say that G represents P if

$$p(x_1, \cdots, x_k) = \prod_{i=1}^{k} P(X_i = x_i | \mathrm{pa}(X_i)), \tag{1}$$

where $\mathrm{pa}(X_i)$ denotes the parent nodes of X_i. (Recall that in a DAG, node Z is a parent of node X iff there is a directed edge going out of Z into X.)

Consider the following DAG

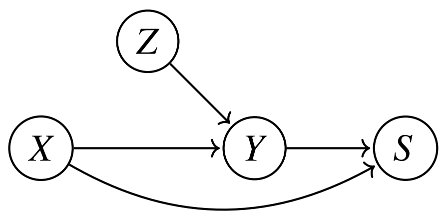

Figure 1: G, a DAG

(a) Write down the joint factorization of $P_{S,X,Y,Z}(s, x, y, z)$ implied by the DAG G shown in Figure 1.

Solution:

$$P_{S,X,Y,Z}(s, x, y, z) = P(X = x)P(Z = z)P(Y = y|X = x, Z = z)P(S = s|X = x, Y = y)\,.$$

(b) Is $S \perp Z \mid Y$?

Solution: No. As a counterexample, consider the case where all nodes represent binary random variables, $P(X = 1) = P(Z = 1) = 0.5$, $Y = X \otimes Z$, and $S = X \otimes Y$, where $\otimes$ is the XOR operator. Then we can see that $S = Z$, whereas knowing Y does not fully determine S or Z.

A version of these solutions from a previous semester erroneously said that this conditional independence did hold. As a result, you may have wrongly heard in section that this statement is true, via faulty algebraic manipulation and/or other algorithms such as the Bayes ball (d-separation). Running these algorithms correctly should show that S and Z are indeed not conditionally independent given Y.

If X is fully removed from G, then we do indeed have $S \perp Z \mid Y$. This is left as an exercise in algebraic manipulation of probability distributions.

(c) Is $S \perp X \mid Y$?

Solution: No. Consider the same example from above with binary random variables. Knowing Y does not determine S, but knowing both X and Y does.

2 Hidden Markov Models: Math Review

A Hidden Markov Model is a Markov Process with unobserved (hidden) states.

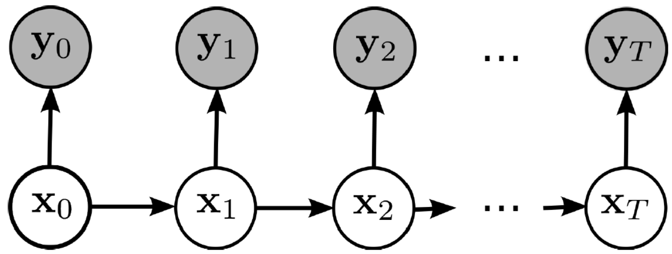

Figure 2: Example Hidden Markov Chain

Consider the following system in $\mathbb{R}^2$, where X_n is the true state at any given time n and Y_n is our observation. Given an initial state X_0, we move to future states by recursively multiplying our current state with transformation matrix A and adding i.i.d. Standard Normal Gaussian noise. When we take an observation Y_n of the true state X_n, we are also exposed to i.i.d. Standard Normal Gaussian Noise.

$$X_{n+1} = AX_n + N(0, I) \tag{2}$$

$$Y_n = X_n + N(0, I) \tag{3}$$

Where we have the $2x2$ transformation matrix A defined as follows:

$$A = \begin{bmatrix} .5 & -.25 \\ -.25 & .75 \end{bmatrix} \tag{4}$$

If we restrict the initial state X_0 to be a unit vector ($\|X_0\|_2 = 1$), determine the following

(a) What are the eigenvalues of A? Is A a positive semi-definite matrix? (Note that $\sqrt{5} = 2.236$)

Solution:

Remember that an eigenvector is a vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$, where the constant λ is the eigenvalue corresponding to $\mathbf{v}$. We manipulate the above equation to be $(A - \lambda I)\mathbf{v} = 0$, which implies that $A - \lambda I$ is a singular matrix since it has an eigenvalue of 0.

$$A - \lambda I = \begin{bmatrix} \frac{1}{2} - \lambda & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} - \lambda \end{bmatrix} \tag{5}$$

We can take the determinant of the above matrix and set it to zero in order for the matrix to be singular, giving us the following characteristic polynomial:

$$0 = (\frac{1}{2} - \lambda)(\frac{3}{4} - \lambda) - (-\frac{1}{4})(-\frac{1}{4}) = \lambda^2 - \frac{5}{4}\lambda + \frac{3}{8} - \frac{1}{16} = \lambda^2 - \frac{5}{4}\lambda + \frac{5}{16} \tag{6}$$

$$\lambda = \frac{1}{2}(\frac{5}{4} \pm \sqrt{\frac{25}{16} - 4(\frac{5}{16})}) = \frac{1}{8}(5 \pm \sqrt{5}) \tag{7}$$

Since $\lambda > 0$ for all possible values, it is a positive-semidefinite matrix (in fact, it is positive definite).

(b) What is the $\| E[Y_\infty] \|_2$? Prove your assertion.

Solution:

Lets look at the first several expressions of the true state X

$$X_1 = AX_0 + N(0, I) \tag{8}$$

$$X_2 = A(AX_0 + N(0, I)) + N(0, I) \tag{9}$$

$$X_3 = A(A(AX_0 + N(0, I)) + N(0, I)) + N(0, I) \tag{10}$$

We note that a particular state can be defined by our original state as follows $X_n = A^n X_0 + \sum_{i=0}^{n-1} A^i N(0, I)$. Thus, our observation of that is $Y_n = A^n X_0 + N(0, I) + \sum_{i=0}^{n-1} A^i N(0, I)$.

Remember that since matrix A is a real symmetric matrix, we can use spectral decomposition to prove that $A^N = (UDU^\top)^N = UD^N U^\top$, where U is a unitary matrix and D is a diagonal matrix of eigenvalues. Note that our eigenvalues are such that $0 < \lambda < 1$. Therefore, $D^N = 0 \Rightarrow A^N = 0$

Thus, when we take expectations and norm, we see that

$$\|\lim_{n\to\infty} E[Y_n]\|_2 = \|E[A^n X_0 + N(0, I) + \sum_{i=0}^{n-1} A^i N(0, I)]\|_2 \tag{11}$$

$$= \|E[N(0, I) + \sum_{i=0}^{n-1} A^i N(0, I)]\|_2 \tag{12}$$

$$= \|0\|_2 \tag{13}$$

$$= 0 \tag{14}$$

(c) Consider the Frobenius Norm of an arbitrary M x N matrix Q, defined as $||Q||_F = \sqrt{\sum_i \sum_j |Q_{i,j}|^2}$, which indicates the "magnitude" or "largeness" of a matrix. Is $||Var[Y_\infty]||_F$ finite or infinite? Prove your assertion.

You may find the following facts to be useful:

(i) Triangle Inequality: $||X+Y|| \leq ||X|| + ||Y||$

(ii) Cauchy Schwarz: $||XY|| \leq ||X||||Y||$

(iii) Geometric Sum: $\sum_{i=0}^{\infty} ar^i = \frac{a}{1-r} \quad \forall r$ s.t. $0 < r < 1; a, r \in \mathbb{R}$

Solution: We will approach this part in the same way as part b). Remember from discussion that for multidimensional i.i.d. random variables X,Y with variance I, and constant matrix B:

$$Var[BX] = BIB^\top = BB^\top \qquad Var[B+X] = Var[X] \qquad Var[X+Y] = Var[X]+Var[Y] = I+I = 2I$$

Therefore, if we examine $\lim_{n\to\infty} Var[Y_n]$, where we define N(0,I) = Q and note that A is symmetric ($A = A^\top$), we see that:

$$\lim_{n\to\infty} Var[Y_n] = Var[A^n X_0 + \sum_{i=0}^{n-1} A^i Q + Q] \tag{15}$$

$$= Var[\sum_{i=0}^{n-1} A^i Q] + Var[Q] = \sum_{i=0}^{n-1} Var[A^i Q] + I \tag{16}$$

$$= \sum_{i=0}^{n-1} A^i I (A^\top)^i + I = \sum_{i=0}^{n-1} (AA^\top)^i + I \tag{17}$$

$$= \sum_{i=0}^{n-1} (A)^{2i} + I = \sum_{i=0}^{n-1} (UDU^\top)^{2i} + I \tag{18}$$

$$= \sum_{i=0}^{n-1} UD^{2i}U^\top + I = U(\sum_{i=0}^{n-1} D^{2i})U^\top + I \tag{19}$$

$$\tag{20}$$

We could stop here and note that $\sum_{i=0}^{n-1} D^{2i}$ is finite since $0 < D_{1,1}, D_{2,2} < 1$. Thus, since D is a diagonal matrix and D^n is also diagonal we can apply the geometric sum formula for each term $\sum_{i=0}^{n-1} (D_{1,1})^{2i}$ and $\sum_{i=0}^{n-1} (D_{2,2})^{2i}$. We then note that the sum is finite, that U and $U^\top$ will preserve magnitude, and I is finite. Therefore, the above limit is finite, which means that the Frobenius

Norm is also finite.

If we want to decompose further, we can use the Triangle Inequality and Cauchy Schwarz Inequality:

$$\|\lim_{n\to\infty} Var[Y_n]\|_F = \|U(\sum_{i=0}^{n-1} D^{2i})U^\top + I\|_F \tag{21}$$

$$\leq \|I\|_F + \|U(\sum_{i=0}^{n-1} D^{2i})U^\top\|_F \tag{22}$$

$$\leq \|I\|_F + \|U\|_F \, \|(\sum_{i=0}^{n-1} D^{2i})\|_F \, \|U^\top\|_F \tag{23}$$

$$\tag{24}$$

We then use the same argument as before to show that the sum of diagonal matrices is a geometric series, and note that I and U are finite matrices. Therefore, both have finite norms and the sum must be finite.