

CS 228 Problem Set 1

Hugh Zhang

January 24, 2017

Problem 1

Prior odds matter when determining if you actually have a disease

$$P(Disease | PositiveTest) = \frac{P(PositiveTest | Disease) * P(Disease)}{P(PositiveTest)}$$
$$.99 * 1/10000 / (9999/10000 * .01 + 1/10000 * .99) = 0.00980392156$$

Problem 2

Since the system is markov, you only need to keep track of the last variable to determine the DP structure.

Let $DP[k][d]$ be

$$\max_{x_1 \dots x_{k-1} \in S^n} P(x_1 \dots x_{k-1}, d)$$

The transition is merely

$$DP[k+1][d] = \max_{v \in S} P(x_{k+1} = d | x_k = v) * DP[k][v]$$

After initializing the base case, you fill in a $m * n$ table with m calculations per transition, which produces the $m^2 n$ runtime as desired.

Problem 3

Imagine a three node graph arranged in a cycle with three directed edges with each node having binary outcomes. Let the probability distribution be such that WLOG if node A is 0, then node child(A) is 1 with probability 1, and vice versa. In this case, since your cycle is of odd length, you come to a contradiction after tracing the path one full cycle.

Problem 4

4.1.1

$$P(A | B, Y) = \frac{P(A, B, Y)}{P(B, Y)}$$

$P(A, B, Y)$ is unfindable because $P(B, Y | A)$ can't be found without conditional independence or knowing $P(Y | A, B)$

4.1.2

Yes, direct application of Bayes Rule

$$P(A | B, Y) = \frac{P(A) * P(B, Y | A)}{P(B, Y)}$$

4.1.3

No, you can't ever find $P(B, Y)$ because everything you have is conditioned on A and you can't ever separate that away.

4.2.1

Yes, because of conditional independence between B and Y given A, you can now do the following

$$P(A | B, Y) = \frac{P(A) * P(B | A) * P(Y | A)}{P(B, Y)}$$

4.2.2

Yes, strictly easier than part 4.1.2

4.2.3

No, you can't ever find $P(B, Y)$ because everything you have is conditioned on A and you can't ever separate that away.

Problem 5

5.1.1

In absence of any other information, A and B are independent given neither is the descendent of the other.

$$.8 * .3 = .24$$

5.1.2

A gives no information on E because E is a nondescendent of A.

Conditioning on the outputs of B:

$$.7 * .1 + .3 * .9 = .34$$

5.2.1

False EHF (V Structure with H known) HFC (Cascade) FCA (Cascade)

5.2.2

False GFC (cascade) FCA (cascade) CAD (common parent) ADB (V structure with D known) DBE (common parent)

5.2.3

False, there is a cascade from B to H.

Problem 6

6.1

0.5

6.2

Drawing the probability density distribution for an admitted student with C on the X axis and I on the Y axis, you get a triangle on the upper right corner with 1/8 the area of the entire unit square.

Since $E[x] = \int P(x) * x$,

$$\int_{.5}^1 8 * (x - .5) * x = \frac{5}{6}$$

6.3

0.5. Baseline nothing changes.

6.4

Given you are admitted and $I = 0.95$, then your creativity ranges from .55 to 1 uniformly. Thus, it is $1.55/2 = 0.775$

You expect this number to be lower than the number for 2, since a higher intelligence requires a less high creativity score on average to be admitted as a student.

Problem 7

7.1

You need a markov blanket on the removed node, which we establish in problem 8 as including the parents, children, and children's parents. The markov blanket isolates all conditional probabilities upon this node.

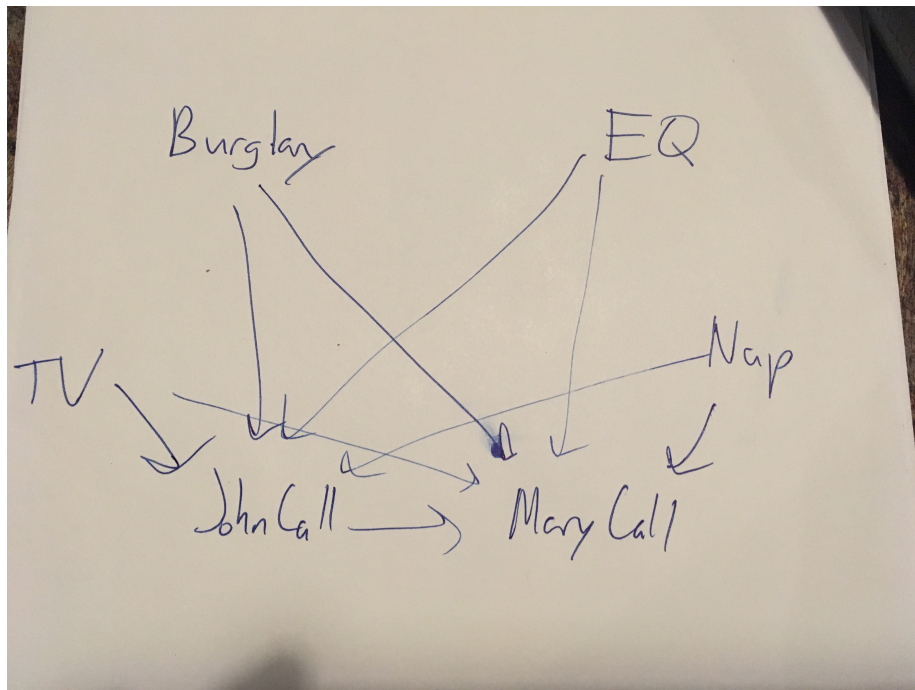


Figure 1: Problem 7 Bayes Net

7.2

Because you can't see the probabilities, you have to just go off of the network structure. This means that any valid network is originally a minimal I map since you can't remove any edge with possibly interfering with the underlying probability distribution.

General gist is if you want to remove node X, for each child of X, add all the parents of X (removing duplicates). Then, topologically sort the children and connect them all to each other in a way that doesn't create a cycle. Then, for each child, draw a line from its parents to all other children in the graph.

So the total update is:

For all children of X: Incoming connections from each parent of X. Incoming connection from each child of X topologically sorted before it. Incoming connection from all parents of any fellow child. (removing duplicates of course).

All other nodes remain the same.

I will now justify the three cases. The intuition is that you are creating a markov blanket. If you are removing X.

Parents of X -> children of X. You need this because cascade allows information to flow.

Children of X to other children of X. Usually, information flows between children of X. But if you don't have these new edges, and condition on all the new parents of X, the common parent information flow is blocked. But this is not true in the old graph. So you need these edges.

Other parents of the children of X to other children of X. This one seems odd, but you seem to need this. If you condition on all the children of X, information

in the old graph flows from all parents to otehr parents based on V strucutres. However, in the new graph without this edge, there is no V structure and the cascade is blocked.

All other nodes should be unaffected because this forms a markov blanket.

Indeed, there are extraneous dependencies, like having the other parents of the children of X always be connected to all the children of X, which is not true on the original graph, but this is OK because an I map only cares that you contain the set of independencies.

If I misread the problem, and you indeed can see the probability distributions for each node X, then you need pick the smallest set S such that

$$P(X | S) = P(X | x1..xn)$$

, then have S be the parents of node X.

Problem 8

8.1

Alright, let's try factoring the graph.

$$\begin{aligned} P(X_i | \text{restofgraph}) &= \frac{P(X_1 \dots X_n)}{\sum_i P(X_1 \dots i \dots X_n)} \\ &= \frac{P(X_i | \text{par}(X_i)) * \prod P(\text{child}(X_i) | \text{par}(\text{child}(X_i))) * \text{restofgraph}}{\sum_{v \in \text{val}(x_i)} (P(X_i = v | \text{par}(X_i)) * \prod P(\text{child}(X_i) | \text{par}(\text{child}(X_i))) * \text{restofgraph})} \\ &= \frac{P(X_i | \text{par}(X_i)) * \prod P(\text{child}(X_i) | \text{par}(\text{child}(X_i)))}{\sum_{v \in \text{val}(x_i)} (P(X_i = v | \text{par}(X_i)) * \prod P(\text{child}(X_i) | \text{par}(\text{child}(X_i))))} \end{aligned}$$

Thus, you can just compute your desired probability by just looking at the probabilities of yourself and your children, and locally normalizing. The rest of the graph cancels out because it has nothing to do with i, and thus you can pull it out of the sum and cancel it out.

8.2

Do the hint exactly, pretty much. Notice that

$$P(A, B) = P(A) * P(B | A)$$

Decompose the Bayesian net into something like

$$P(A) * P(B) * P(C | A, B) \dots$$

using a topological sort.

So the full equation is:

$$\prod_{i=1}^n P(x_i | \text{par}(x_i))$$

Then go down and generate the net by first generating the parents, then their children using their values, until you have generated the whole net.

Problem 9

9.1

$$2^{28 \times 28}$$

9.2

$$2^{28 \times 28} - 1$$

Requires one less than 9.1 because you need a probability for every vector, but you know the distribution sums to one, so you can save one parameter.

9.3

Z can take values from $[-3, 3]$ stepping at .25 intervals, which implies $[-12, 12]$ at integer intervals = 25 different values. Thus, to specify each of Z_1, Z_2 , you need $(25 - 1) * 2 = 48$ parameters

To specify each of $X_1 \dots X_{784}$, you need $P(0 \mid Z_1, Z_2)$ for all Z_1, Z_2 since $P(1 \mid Z_1, Z_2) = 1 - P(0 \mid Z_1, Z_2)$. Thus, you need $25 * 25 * 28 * 28 = 490000$. Thus the answer is $490000 + 48 = 490048$

9.4

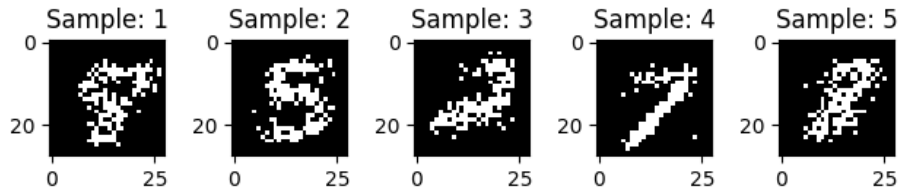


Figure 2: Sampled images

9.5

It seems that the x axis controls whether you are round or linish, while the y axis controls the degree / number of attributes.

9.6

9.7

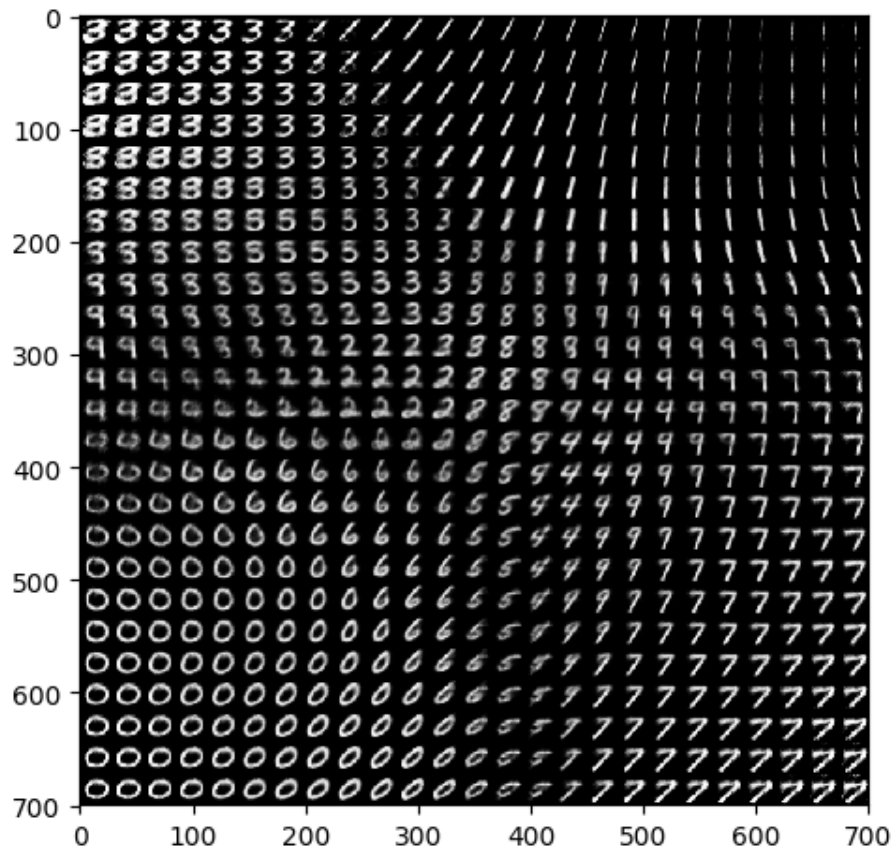


Figure 3: Role of Z_1 and Z_2

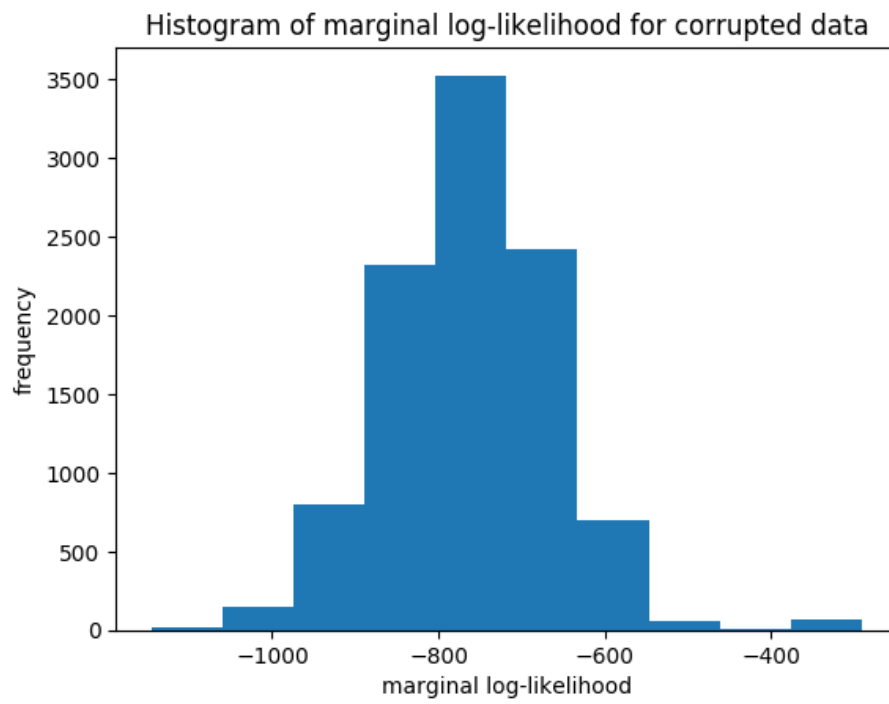


Figure 4: Corrupted Test Images

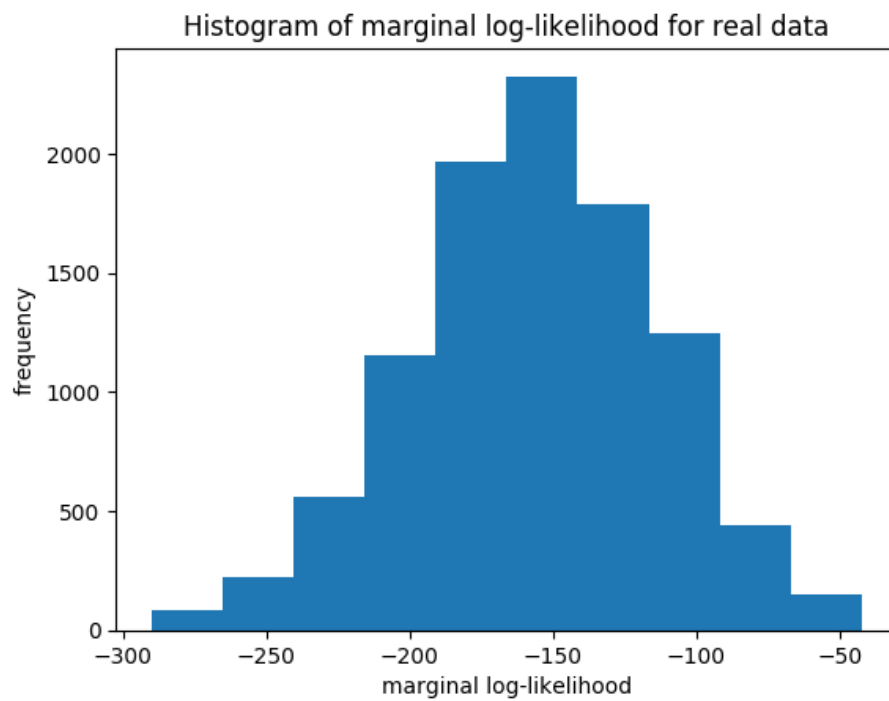


Figure 5: Valid Test Images

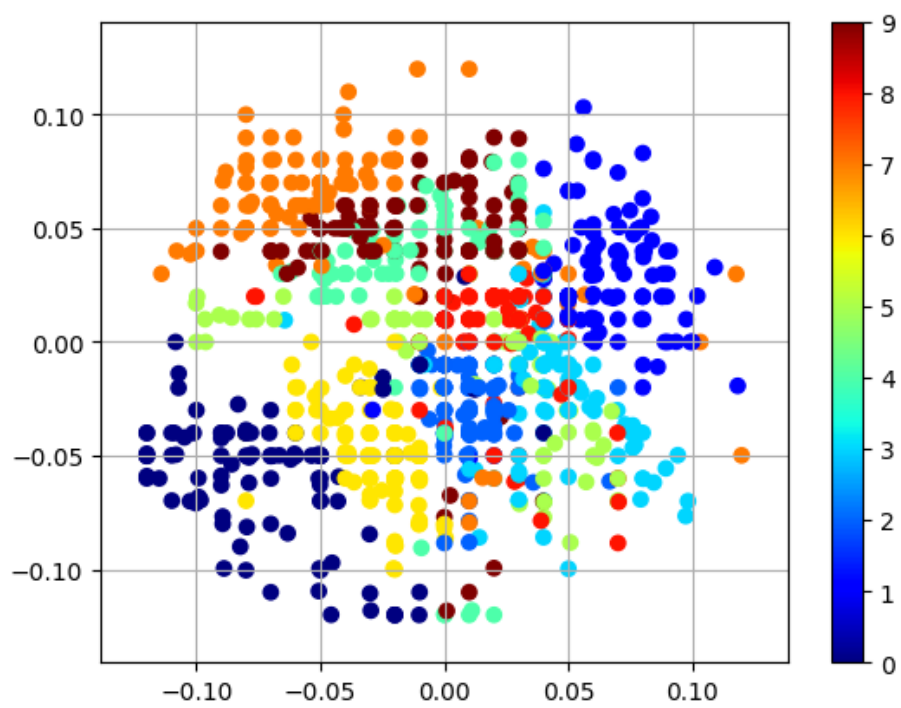


Figure 6: Scatter plot