

Problem Set #2

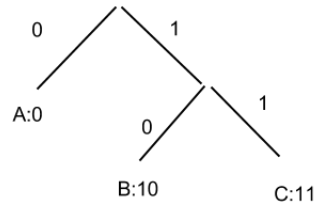
Matthew Kirk

November 24, 2020

1 Encoding of A,B,C

You have to communicate a signal in a language that has 3 symbols A, B and C. The probability of observing A is 50% while that of observing B and C is 25% each. Design an appropriate encoding for this language. What is the entropy of this signal in bits?

Since we know that the probability of each of these symbols showing up is not uniform we can build a variable width encoding. To do that we will first build a tree which assigns less bits to the higher probability 'A' and then two bits to 'B' and 'C'. It would look like this.



That means that A would be '0', B would be '10' and C would be '11'. This would make the average size 1.5 bits long $(0.5 + 0.25 * 2 + 0.25 * 2)$. In this case that would also be the entropy of this language since it is the most compact bits one can transfer while not losing any information.

2 Comparison of Genetic Algorithms, Simulated Annealing, Randomized Hill Climbing, and Mimic

2.1 What problems are each suited for? And what are their strengths and weaknesses?

1. Genetic algorithms are exceptional at getting out of local optima. The entire idea of introducing crossover and mutation over generations will avoid many problems that other algorithms like hill climbing and mimic would fall into. They are well suited for problems that are highly noisy

in terms of having lots of local optima. A good problem that it is well suited for is finding optimal neural network weights, which can be a noisy process.

2. Simulated annealing are excellent at solving a noisy hill climbing algorithm. They can get stuck like any other algorithm because over time the temperature cools meaning that it stops exploring and starts exploiting therefore becoming more of a hill climber. They are excellent at finding solutions quickly though and do wonders on data that have a very high max. Simulated annealing also operates extremely fast so therefore you can center in on something quickly. Although that fastness doesn't necessarily mean that it will converge as fast as other algorithms. A problem that simulated annealing is particularly well suited for is something like a 4 peak problem. It does very well at exploring and then cooling down at the end to center in on a global maxima.
3. Randomized hill climbing is a canonical example of optimization. It has some downsides though which is that it tends to get stuck a lot which is why randomized restarts happen. If it was given a piece of data that was somewhat noisy but still had quite a smooth surface to it to optimize it would do well. The one thing it has for it though is that it's extremely fast to compute. It operates very fast. Randomized hill climbing is well suited for the 4 peak problem as well. It is kind of the cousin of simulated annealing in a way.
4. MIMIC is really a great algorithm, it finds structure of the data as it samples through space and makes better guesses on that. It will converge on a solution in orders of magnitude less iterations. Though all of that comes at a hefty hefty cost. While it might take 100 iterations to get to a solution those 100 iterations might be dog slow. It is well suited for data that has structure underneath it. MIMIC I would say is well suited for something like the k-color problem.

3 Show that the K-means procedure can be viewed as a special case of the EM algorithm applied to an appropriate mixture of Gaussian densities model.

K-Means is really a special case of the EM algorithm because unlike EM which softens the boundaries between clusters K-Means has a hardened line. In practice this would mean that all of the hidden variables $Z_n \in \{0, 1\}$. There wouldn't be any sort of softening or probabilistic definition of clusters.

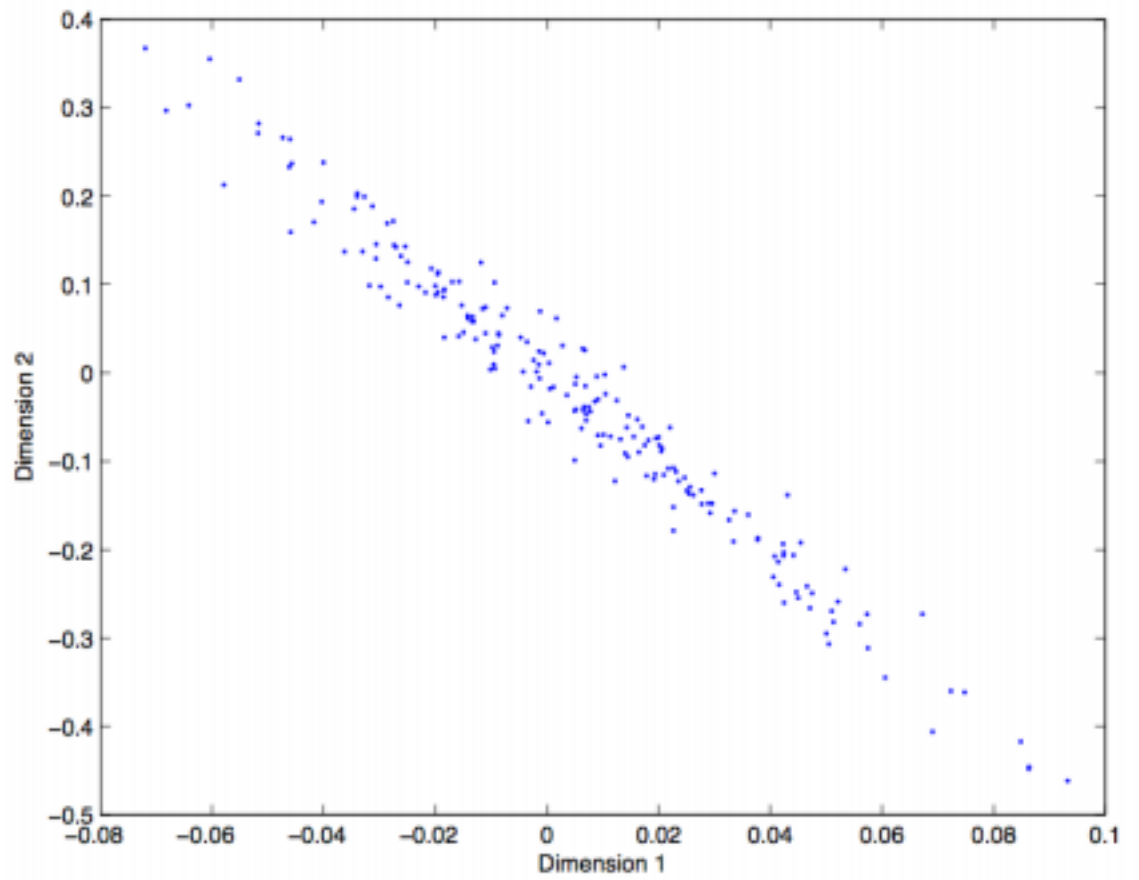
When we harden up the underlining hidden variables then the EM clustering algorithm becomes something more like:

$$E[Z_{ij}] = \frac{P(x = x_i | \mu = \mu_j)}{\sum_i^k P(x = x_i | \mu = \mu_j)} \Leftrightarrow \mu_j = \frac{\sum_i E[Z_{ij}] x_i}{\sum_i E[Z_{ij}]} \Leftrightarrow \mu_j \in \{0, 1\} \Leftrightarrow \mu_j = \frac{\sum_i x_i}{|C|}$$

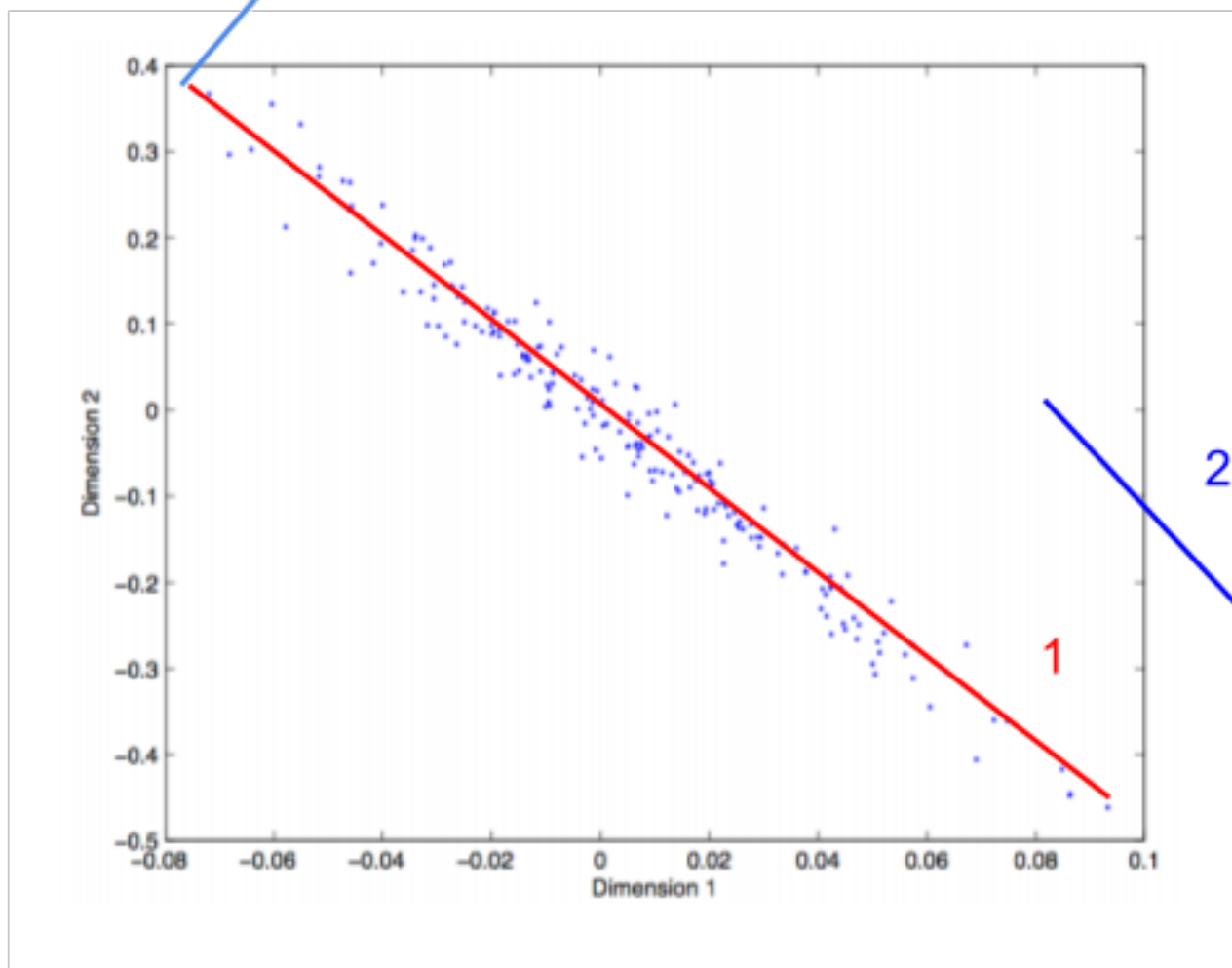
Basically since we are limiting Z_{ij} to be either 0, or 1 then the expectation of Z_{ij} will be either 0, or 1. The expectation step is really just assigning data points to each individual cluster denoted by j . The maximization step will end up calculating the means of the data because Z_{ij} is no longer between 0 and 1 it is either 0 or 1 that means that it will end up being purely the mean.

4 Plot the direction of the PCA components in the figures given

The original graphic



Annotated with PCA graphs.



In the first figure first PCA direction is fairly easy to detect. It is the one with the maximal variance and therefore is looking downwards. The second component is just orthogonal to that.

The second figure was tougher to detect. I am guessing that since it looks like the positive direction has the highest variance that that is where the first principal component is. And likewise the orthogonal direction from that should be the second principal component.

5 Match the clustering method with the figure

Which clustering method(s) is most likely to produce the following results at $k = 2$? Choose the most likely method(s) and briefly explain why it/they will work better where others will not in at most 3 sentences. Here are the five clustering methods you can choose from: - Hierarchical clustering with single link - Hierarchical clustering with complete link - Hierarchical clustering with average link - K-means - EM

I would suspect that this is single link clustering. The reason for that is the moat in the middle is bigger than the points next to each other. So if you pick a random two points in there (Assuming you get one on each side of course) then it will iterate through that line until it finishes with two clusters. It couldn't be K-Means because K-Means would end up averaging the two and most likely split the graphic vertically instead of horizontally. EM Clustering might pick this up but I think it would have a closer answer to K-Means than this.

This is definitely the work of K-Means. The two clusters are spherical in nature (a characteristic of K-Means) and center around what looks like centroids. EM clustering would probably introduce some soft layers in here so most likely it is not EM clustering. Single Link clustering would futz up the middle too much for this to happen.

I would suspect this is EM clustering. The reason for that is the soft clustering in the middle. KMeans would most likely split this data into two pieces vertically while single link clustering would come up with something completely different. EM clustering is the only algorithm I could see that could come up with something so soft like this.

6 The Haunted House

1. To formulate this haunted house as a MDP we need a few things defined: state set, action set, state transition matrix, and a reward function.

The haunted house can be in the following states: * Laughter * Quiet

The actions one can take are: * Play the organ / Don't play the organ *

Burn incense / Don't burn incense

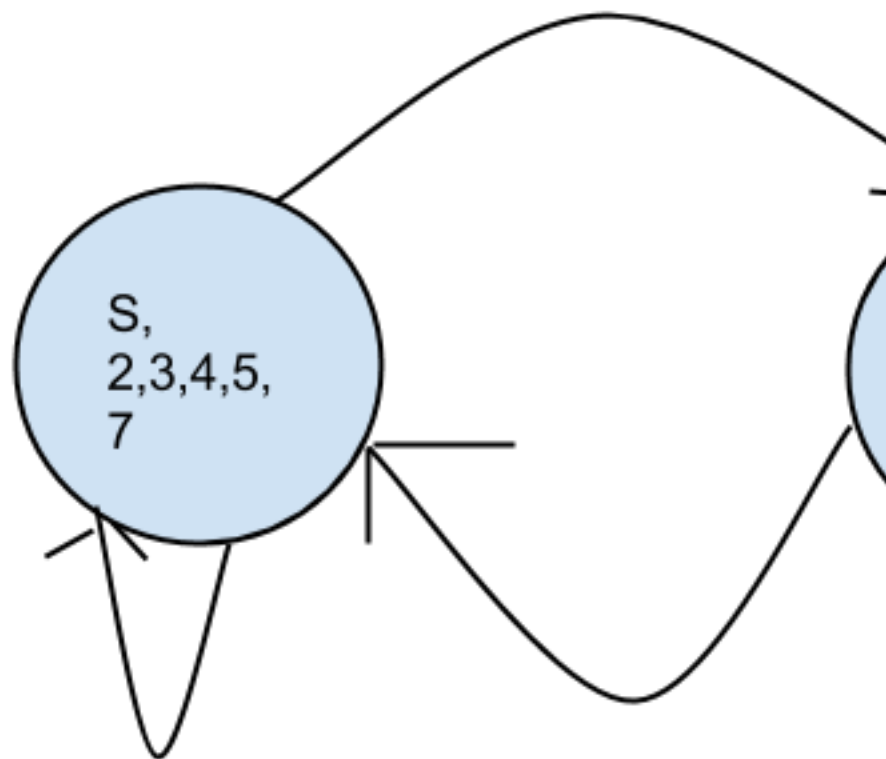
State transitions will be contained in this:

| State | Action | New State | Probability |
|-----------|-------------------------------|-----------|-------------|
| Laughter | Play organ | Quiet | 100% |
| Not Quiet | Playing Organ | Not Quiet | 100% |
| Quiet | Burn Incense don't play organ | Quiet | 100% |

7 Bellman Equations

8 Grid World

The MDP of this looks something like this.



There are three groups of states basically. You can travel from (S,2,3,4,5,7) to (6,8) to (G). The other thing you can do is stay in the (S,2,3,4,5,7) group as well.

Computing the value function for this would be:

Iteration 0:

For $U_0(S) = 0$ We set all utilities to equal 0 in the first iteration. That is mostly for convenience and to set it equal to something.

| State | Utility |
|-------|---------|
| S | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| 5 | 0 |
| 6 | 0 |
| 7 | 0 |
| 8 | 0 |
| G | 10 |

Iteration 1:

The only states that are relevant are 8 and 6 since the rest will be zero for now. In cases of 8 and 6 we can calculate

$$U_1(8,6) = 0 + 0.8 * (\max_a \sum U_0(S))U_1(8,6) = 0.8 * 10 = 8$$

| State | Utility |
|-------|---------|
| S | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| 5 | 0 |
| 6 | 8 |
| 7 | 0 |
| 8 | 8 |
| G | 10 |

Iteration 2:

This is where things become more interesting. We have 3,5,7 which are close to 8,6 now which do have utility so therefore we can now calculate values for them. So

$$U_2(3) = 0.8 * (0 + 8) = 6.4 = U_2(7)$$

$$U_2(5) = 0.8 * (8 + 8 + 0 + 0) = 12.8$$

$$U_2(8) = 0.8 * (10 + 0) = 8$$

| State | Utility |
|-------|---------|
| S | 0 |
| 2 | 0 |
| 3 | 6.4 |
| 4 | 0 |
| 5 | 12.8 |
| 6 | 8 |
| 7 | 6.4 |
| 8 | 8 |
| G | 10 |

Iteration 3:

Calculating out all variables now:

$$U_3(2, 4) = 0.8 * (6.4 + 12.8 + 0) = 15.36 \quad U_3(3, 7) = 0.8 * (8 + 0) = 6.4$$

$$U_3(5) = 0.8 * (0 + 8 + 8) = 12.8 \quad U_3(8, 6) = 0.8 * (12.8 + 10 + 6.4) = 23.36$$

It's interesting to see this spread but eventually it'll converge on a good solution that directs traffic towards G.

9 Nash Equilibrium

| | | | |
|--------|---|-----|-----|
| Game 1 | | A | B |
| | A | 2,1 | 0,0 |
| | B | 0,0 | 1,2 |

This is the battle of the sexes game. There are actually two nash equilibria in this case. Either both players chose A or both Players chose B. That is because if one goes with A and the other goes with B then they will both get 0 which is not the best outcome.

Game 2

| | | |
|---|-----|-----|
| | A | B |
| A | 2,1 | 1,2 |
| B | 1,2 | 2,1 |

In this game if Player 1 picks the strategy A then player 2 would pick B to maximize their return. If player 1 picks strategy B then Player 2 will pick A. In both of these cases the player ends up with a 1 payoff while player 2 gets 2.

Now in player 2's case. If they pick B then player 1 will pick B meaning a 2,1 payoff. If player 2 picks A then player 1 will pick A as well.

That means all four are nash equilibria... Given each strategy each player is playing you cannot do any better than to respond in given fashion.

| | | | |
|--------|---|-----|-----|
| | | A | B |
| Game 3 | A | 2,2 | 0,0 |
| | B | 0,0 | 1,1 |

The nash equilibrium in this case is for all players to chose A all times. You cannot do any better than that and it strictly dominates all other strategies.