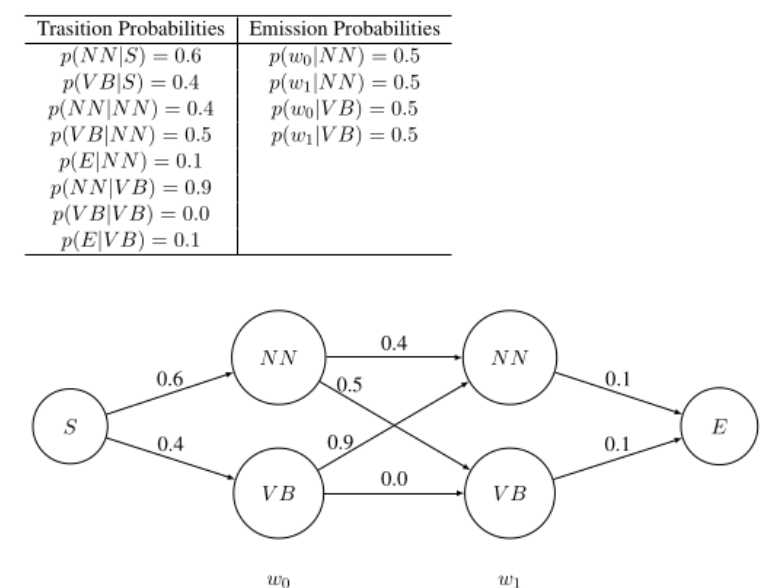
Q1. In a few sentences, what makes natural language processing difficult?  
**Ans**: Language is inherently ambiguous, with word and Phrases often having multiple meanings depending in context. This can result in inaccurate analysis or miscommunication. Key factors that  NLP difficult:  
Ambiguity, Semantic understanding, Linguistic diversity, Context dependence  
  
Q2. Given the following HMM model, predict the most probable part-of-speech tag sequence for the sentence “w0 w1" and give it's probability.  Be sure to include emission and transition probabilities.  Do not include emission probabilities for start (S) and end (E).  
  
  
**Ans:**  
T = 0:  
 δ0​(NN) = P(NN)\* P(w0|NN) => 0.6\*0.5 = 0.3  
 δ0​(VB) = P(VB)\* P(w0|VB) = > 0.4\*0.5 = 0.2  
T = 1:

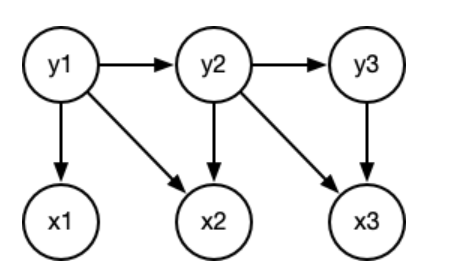
δ1(NN) = max[δ0​(NN)\*P(NN|NN) , δ0​(VB)\*P(NN|VB)]\*P(w1|NN)  
 = max[0.3\*0.4 , 0.2\*0.9]\* 0.5  
 = [0.12 , 0.18]\*0.5 = [0.06,0.09]

δ1​(VB) =max[δ0​(NN)\*P(VB|NN) , δ0​(VB)\*P(VB|VB)]\*P(w1|VB)  
 = max[0.3\*0.5 , 0.2\*0.0]\* 0.5  
 = [0.15 , 0.0]\*0.5 = [0.075,0.0]

Ans = P\* = max[δ1(NN), δ1​(VB)] = 0.09   
 [NN,NN]

Q3.   
1. In Bayesian learning, what is p(θ) in the equation P(θ|D) = P(D|θ) P(θ)  
2. What does the MAP estimate maximize?  
3. What does the MLE estimate maximize?  
**Ans:**  
1. p(θ) : called the prior, the prior probability mass function of the hypothesis.  
2.MAP : If we estimate θ by maximizing p(θ|D), θ is the value of the hypothesis.  
3.MLE: If we estimate θ by maximizing p(D|θ), θ is the value of the hypothesis.

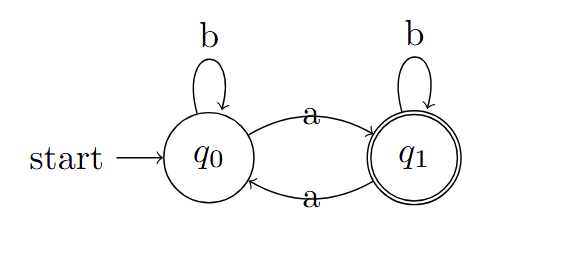
Q4.1) Write down the joint probability for the following Bayesian network.

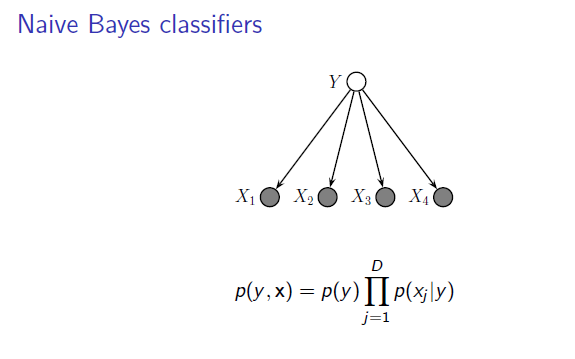
  
2) How would you sample from the joint probability of this Bayesian network?  What order of random variables would you pick?  
**Ans:**  
1.  
Node: X1,X2,X3,Y1,Y2,Y3  
Parent Node = Y1  
X1 = P(X1|Y1)  
Y2 = P(Y2|Y1)  
X2 = P(X2|Y1,Y2)  
Y3 = P(Y3|Y2)  
X3 = P(X3|Y2,Y3)  
P(X1,X2,X3,Y1,Y2,Y3) = P(Y1)\* P(X1|Y1)\* P(Y2|Y1)\* P(X2|Y1,Y2)\* P(Y3|Y2)\* P(X3|Y2,Y3)  
2. According to the structure of Bayesian network start with variables without parent node and gradually move to the node with dependencies.

Q5: What types of ambiguities does the following sentence have? How many possible readings  
can you find for it? Explain what each reading is.

I saw a cat on the roof.

**Ans :**It show structural ambiguity in "*on the roof*"  
**Reading 1 (Easy):** A cat that was physically located on the roof.

**Reading 2:** The speaker saw the roof and a cat is on the roof. (The cat was not the primary focus of the speaker's attention, but was simply somewhere on the roof while they were looking at it).  
Q6.   
For the NFA defined by   
Q = {q0, q1}   
Σ = {a,b}   
F = {q1}   
δ (q0, a) = q1,   
δ (q0, b) = q0,   
δ (q1, a) = q0,   
δ (q1, b) = q1  
Draw the NFA.  What best describes the strings it accepts?  
**Ans:**  
We can know the NFA can accepts strings that end with 𝑎.  
Q7: In your own words, what is a morpheme?  
**Ans**: Morpheme: The Smallest meaning units in any language.

Q8: Give an example of a bound morpheme in English.  
**Ans:**  
 1. Un- : Unhappy  
 2. -ed : played  
Q9: For each word, match it with the correct morphological phenomena.  
**Ans:**  
 It’s : Cliticization  
 brushless : Compounding  
 Cleaning : Inflection  
 accomplishment : Derivation  
Q10: For a Naive-Bayes classifier, what is the form of the joint probability of the data p(x,y)?  Draw the Baysian network for it.  
**Ans:**  
 p(x,y) = p(y) \* Π(p(x\_j| y)  


Q11: In your own words, what is the difference between Bayesian and Frequentist learning?  
**Ans:**  
**Frequentist :** Assigns probabilities to data, not to hypotheses, focuses on the data collected, without incorporating prior beliefs  
**Bayesian:** Approach assigns probabilities to hypotheses, and  incorporate prior knowledge into the analysis, updating hypotheses probabilities as more data become available.  
  
Q12. Here are some grammatical and ungrammatical sentences (\* indicates an ungrammatical sentence):

The mouse jumped over the cat.

Over the cat, the mouse jumped.

\* Over the, the mouse jumped cat.

According to linguistic tests for constituency, what does this indicate about the words “over the cat” and "over the" in this sentence?  Which test(s) did you use to make your determination?

**Ans:** The original meaning is The mouse jumped over the cat.  
But "over the cat"  also mean mouse jumped over the cat ,but it highlight over the cat that behavior and "Over the" mean come the meaning is total different compare to the original meaning.  
we can used syntax to determinate the sentence id grammatical or ungrammatical   
also we can used the context to confirm grammatical correctness

Q13. Is it possible to use a multiclass probabilistic classifier to build a language model?  If so, describe how you would train it, and how you would compute the probability of a sequence of tokens.  Use words and/or equations.  Which probability equation, if any, are you using (i.e. Bayes rule, the chain rule, or marginalization, etc)?

**Ans :** Yes, we can used the multiclass probabilistic classifier to build a language model . To process the data we can used the n-gram format to get the feature, and the classifier I want to choose the Naive Bayes model used chain rule to calculate the probability of a sequence of tokens, and used the marginalization to Marginalize part of the context.