Net Id: RPK9907

Free-response questions (8 points)

1. Comparing Naive Bayes with and without unlabelled data (1 points)

See `src.free_response` for code that will help you with part (a) and (b).

Create a dataset using the provided `src.data.build_dataset` function with args `n_docs=100`, `max_words=2000`, and `vocab_size=1000`. You do not need to split this dataset into train and test splits. Instead, you will fit both your `NaiveBayes` and `NaiveBayesEM` models on the `data` matrix and `labels` array returned by this function. For (a) and (b) below, evaluate your models on the entire dataset and add your results to this table.

Model	Accuracy	Log Likelihood
NB	98.27586	-504019.4819
NB + EM	68.96551	-846701.2475

(a) Use each fit model to predict labels for each labelled (post-1964) speech. Calculate the accuracy of each model's predictions and include them in the table above. Unlabelled speeches (`labels == np.nan`) should not factor into this accuracy calculation for either model.

(b) Calculate the log likelihood of the entire dataset according to each model and include it in the table above. unlabelled examples should not contribute to the likelihood of the non-EM model.

(c) Discuss the differences in accuracy and likelihood between the `NaiveBayes` and `NaiveBayesEM` models. Provide an explanation for why the model with the higher accuracy performs better. Similarly, explain why one model has a higher likelihood than the other.
Ans. It can be concluded from the above table that Naïve Bayes is model as it has greater accuracy and log likelihood than that of the Naïve Bayes EM.

Lowest:

jobs	-0.003454322
work	-0.003089014
every	-0.002990027
years	-0.002695165
thats	-0.002631741

Highest

free	2.49E-03
government	0.002720735
america	2.73E-03
world	2.93E-03
freedom	3.02E-03
federal	3.07E-03

2. Naive Bayes and probabilistic predictions (3 points)

For these questions, use the same dataset as in the previous question. See `src.free_response` for code that will help you with part (b).

(a) Define `f(x) = p(x | y = 1) - p(x | y = 0)` as the difference in "party affiliation" of a word `x`, which takes positive values for words that are more likely in speeches by Republican presidents and takes negative values for words more likely in speeches by Democratic presidents. Compute `f(x)` for each word `x` in the vocabulary. What are the five words with the highest (positive) and lowest (negative) values of `f(x)`?
Ans.

(b) Use the code in `src.free_response` to look at the probabilities output by the NaiveBayesEM model both when it makes a correct prediction and when it makes an incorrect prediction on the labelled speeches. We can think about these probabilities as describing the "confidence" of the classifier -- if it outputs a probability of 50% for both labels, the model is unconfident. When the model's probability of the predicted label is close to 100%, it is very confident in its prediction. What do you notice about the confidence of your NaiveBayesEM classifier, both when it is correct and when it is incorrect? What explains this behaviour?
Ans. Naïve bayes EM classifier acts as too confident while prediction while classifying correct and incorrect. The reason for this is that there are some problems in Naive Bayes EM. We consider that the all the words are conditionally independent, but this are rarely true. But this does not affect classification accuracy however this point makes it too confident in prediction. We can observe that the probabilities estimated by Naïve Bayes are skewed towards 1 or are very close to 1. Hence as a result it makes the output very skewed in certain area. This as a result makes the Naïve Bayes EM over confident.

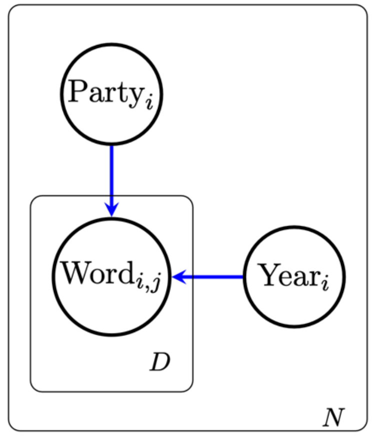
While the data is more this effect gets compounded and hence it becomes more confident.

(c) Suppose we were using a machine learning classifier in a high-stakes domain such as predicting clinical diagnoses of patients in intensive care. What might be a danger of having a classifier that is confident even when incorrect?
Ans. If a classifier is used in clinical diagnosis makes overconfident decision then it may lead to:

Then such classifier may predict wrongly which can some time tell patients that they are diagnosed with a disease who are not infected or vice- versa. It may lead to death of patient. Using such classifier in real world clinical system it can lead to many different social, economic and health crisis.

(d) What is one thing we could do to change our NaiveBayesEM classifier to cause it to make less confident predictions.
Ans.

3. Modelling the year of speeches (1 point)
In addition to our label that represents the political party of the speaker, we could also include a variable for the year in which the speech was given. Suppose we now choose to use following graphical model, where `i` indexes into the `N` documents and `j` indexes into the `D` words in our vocabulary.



(a) What are the (conditional) independences in this model? Consider the variables `Year_i`, `Party_i`, `Word_i, j`, and `Word_i, k` where `j != k`.
Ans. The given situation in the question is a head-to-head type of graph where word_i_j act as a collider.

- party_i and year_i are independent without conditioning as word_i_j acts as a collider.
- if we condition on party_i, then word_i_j and year_i will become independent.
- if we condition on year_i, then word_i_j and party_i will become independent.
- all the word are independent of each other i.e., Word_i_j is also independent of Word_i_k.

(b) Remember that our standard Naive Bayes model with binary labels, we needed `2 * D + 1` parameters to represent the joint distribution, where `D` is the number of words in the vocabulary. Assuming that Year is a discrete variable with `Z` possible values, how many parameters do we need in our conditional probability tables to represent the distribution that includes Year? Explain your answer.

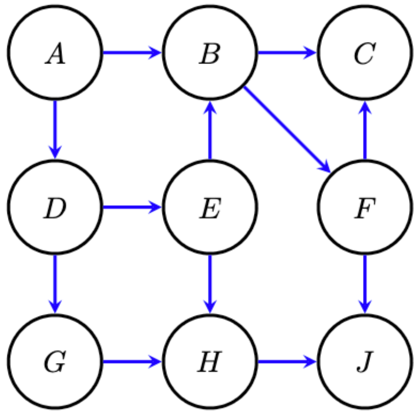
Ans. P (word_1, word_2, …, word_d, year, party)
P (word_1, word_2, …, word_d| year, party) *P (year, party)
P (word_1|year, party) * P (word_2|year, party) * … * P (word_d|year, party) *P(year)*P(party)

1 parameter: P(party).
(Z-1) parameters: P(year)
(d-1) parameters: each P (word_j|year, party)
2z combinations of year, party.
2z(d-1) parameters: problem.

So total parameters = 2z(d-1) + z-1+1 = z(2d-1)

4. D-Separation and conditional independence (3 points)

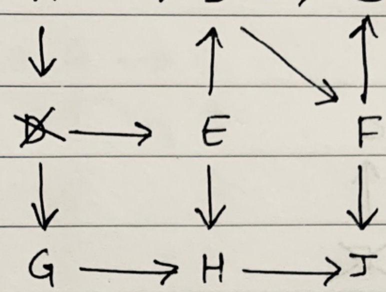
Consider the directed acyclic graph below. We will write "Is A ⊥ B | C?" to mean "Is A conditionally independent of B given C?" For each question, provide your explanation in terms of blocked and unblocked paths.
For example, if the question were "Is A ⊥ G | D?", the answer "Yes" would not be worth points. You must say: > Yes, because any path through A->D->G is blocked at D, and any path through E->H<-G is blocked at H, and all paths from A to G go through either D or H.



(a) Is A ⊥ E | D? Why?

Ans.

(a) $A \perp E \mid D$?



① $A \rightarrow \boxed{B} \leftarrow E$

acts as a collider

② $A \rightarrow B \rightarrow \boxed{C} \leftarrow F \rightarrow \boxed{J} \leftarrow H \leftarrow E$

acts as a collider

③ $A \rightarrow \cancel{D} \rightarrow E$

D is blocked.

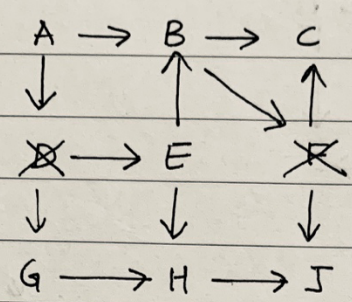
∴ There is No direct connection between A to E.

∴ Yes, A is conditionally independent of E given D

(b) Is A ⊥ E | D, F? Why?

Ans.

(b) $A \perp E \mid D, F$?



① $A \rightarrow \cancel{D} \rightarrow E$ because D blocks the path.

② $A \rightarrow B \leftarrow E$ because B acts as collider

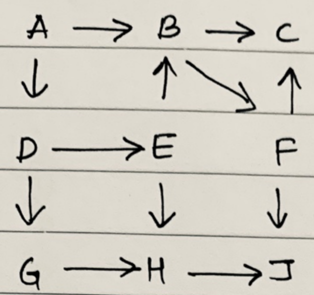
③ $A \rightarrow B \rightarrow C \leftarrow \cancel{F} \rightarrow J \leftarrow H \leftarrow E$ because F is blocked & J acts as collider

∴ As there are no paths available to reach from A to E, A is ~~independently~~ conditionally independent of E given D & F.

(c) Is A ⊥ B | D, E? Why?

Ans.

(c) $A \perp B \mid D, E$?

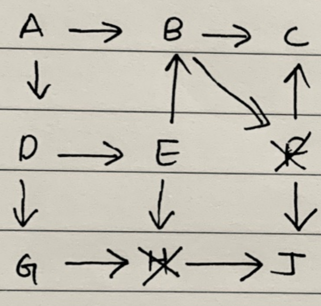


A is directly connected to B. Therefore No, A is not conditionally independent of B given D & E.

(d) Is A ⊥ J | F, H? Why?

Ans.

(d) A ⊥ J | F, H?



① A → B → ~~F~~ → J

because F is blocked.

② A → $\boxed{B}$ ← E → ~~H~~ → J

because H is blocked & B acts as collider.

③ A → D → ~~B~~ G → ~~H~~ → J

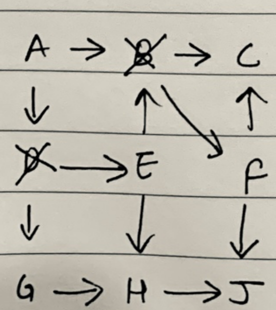
because H is blocked.

∴ yes, A and J are conditionally independent given F, H.

(e) Is A ⊥ J | B, D? Why?

Ans.

(e) A ⊥ J | B, D?



① A → ~~B~~ → C ← F → J

because B is blocked & C is collider

② A → B ← E → H → J

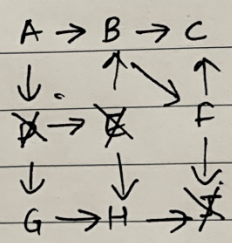
can pass as a collider under condition allows ~~for~~ it to unblock the path.

∴ NO, A ~~& J~~ ~~are~~ is not conditionally independent of J given B & D.

(f) Is A ⊥ H | E, D, J? Why?

Ans.

(f)



① $A \to B \to C \leftarrow F \to \cancel{J} \leftarrow H$

can not pass as even collider under condition unblocks path, there is another collider C that blocks it

② $A \to B \leftarrow \cancel{E} \to H$ cannot pass

③ $A \to \cancel{D} \to G \to H$ cannot pass.

$A \perp H \mid E, D, J$?

∴ Yes, A ~~are~~ is conditionally independent of H given E, D, J