Name:

NetID:

Note: This quiz will be a set of practice problems for the upcoming exam. Complete this at your leisure outside of class but please bring in an **INDIVIDUAL** copy of the quiz **ON PAPER** to turn in at the start of the exam. While doing the work feel free to work with your classmates, but I'd like you all to turn in a copy individually. You don't necessarily need to print and fill out this exact document, just writing the answers on your own piece of paper will be fine.

1.) What is the bias/variance tradeoff?

Bias is the implicit assumptions we make in choosing the model

Variance is how our model responds to seeing different data

↑Bias is usually the model is too simple and underfits

↑Variance is overfitting and not robust

2.) How does k-Nearest Neighbors classify points when used for classification?

It uses a voting method where you find the k nearest point then pick the class most of them are

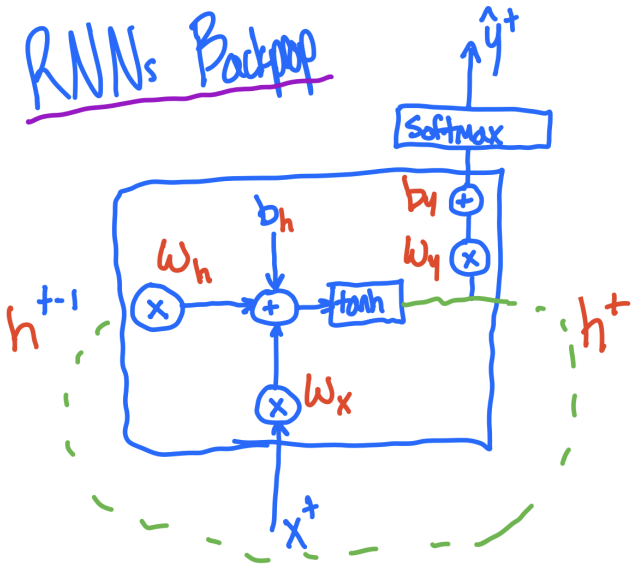
3.) What is the purpose of the Gaussian distribution in a Naive Bayes classifier?

It's the distribution we fit to our underlying data and used to estimate the likelihood of a new sample belonging to a certain class. Assumes our data's distribution is gaussian

4.) What is polynomial feature expansion?

It's when we take a feature and essentially create new synthetic features out of the orginal one by raising it to increasing powers so our model will hopefully find new patterns

5.) Given the network diagram, draw the computation graph for the network for the sentence “Good Luck” with a word level tokenizer, only producing output for the final token. (5 pts.)



W_x X="Good" h_0 W_h b_h

$z_1 = W_x X + h_0 W_h + b_h$

$h_1 = \tanh(z_1)$

X="luck"

$z_2 = W_x X + h_1 W_h + b_h$

$h_2 = \tanh(z_2)$

$A^t = W_y h_2 + b_y$

$y^t = \text{Softmax}(A^t)$

6.) What are tokens and embeddings in relation to NLP tasks?

tokens are a mapping of a word (or maybe smaller) to a representative number. Embeddings are like word feature vectors we can retrieve with th token.

7.) Convolutional Networks use filters or kernels to process images. What do these look like and what purpose do they serve?

They are small grids of pixels that we "slide" over th image that fire if th part of th image we're checking matches th filter

8.) What are the two types of pooling layers? Why do we use them?

Max & Avg pooling are used to downsample or "blur" images to help filters match

9.) What is the difference between Djikstra's and A*?

A* uses a heuristic to help "direct" itself to the goal state

10.) We say Attention(Q, K, V) is a “dictionary lookup”, explain this metaphor.

Q: Query, what X looks for
K: Key, what X has
V: Value, Context X provides

QK gives us our attention weights which we (QK) V to the find the context to add to our input

11.) What is layer normalization and why do we use it in a transformer?

It's a way of setting 0 mean, 1 STD to help stabilize the traing process and prevent vanishing/exploding gradients

12.) How are minimax and alpha/beta pruning related? Do we still have to use minimax if we use alpha/beta pruning?

a/b pruning is an improvement on top of minimax to help prune branches we know won't help but a/b pruning still runs on top of minimax

13.) What are the differences between gradient descent, stochastic gradient descent, and mini-batch stochastic gradient descent?

GD: is updating your weights after every sample sequentially

SGD: th same as above but with a randomly selected sample

MB-SGD: same as SGD but you do a batch of samples ad average th gradients

best!

14.) What is a linear transformation and how does it relate to deep learning?

A linear trasformation is like using a matrix of operations to transform or project a point from one space to another. Basically all of DL is a system for learning the transform matrix

Bonus.) The draft website for next semester is available at nextai.williamtheisen.com I'd love it if you were willing to look at it and give me any thoughts you may have. Our TA, Tom, suggested that if you turned in the exam practice packet you could get points on the exam which I thought was a really good idea!

6/10 let Tom teach the whole thing