18.25
18

Wednesday, April 19, 2023 5:49 PM

1a: Vector multiplication axioms include the same sorts of things:

1.5/2

1. Associative Property:
   a(bu) = (ab)u
2. Distributive Property:
   a(u+v) = au + av
   (a+b)u = au + bu
3. Scalar Identity
   1(u) = u

forgot additive inverse

$-V+V=0$

1/1

1b: A field. This is because the definition of a field is a set where addition, subtraction, multiplication, and so on are defined like with regular real and rational numbers. This is commonly applied to scalars in vector space within linear algebra as all of those functions apply and you can even treat any real number as a scalar and even try to consider a scalar as a certain type of vector in some contexts.

1.75/2

2a: The n+1 dimensional neuron has an initial bias that must be overcome which isn't variable. The inputs and outputs are still the same for both hyperplanes. The only difference is that you have an X0 term to multiply the W0 term with. However, since the X0 term is ***always*** set to 1, it is the same as the bias by the rule of identity, therefore, there isn't any difference between the two hyperplanes, besides that H includes the weight inside the weighted sum and H underscore incorporates it via the inequality.

write out the math

1/1

2b: The special property is that it passes through the origin of the n+1 dimensional space.

2/2

2c: In the first case, we haven't changed what the hyperplane defines, just scaled it. The firing rule doesn't change for H underscore, because W0 is scaled and -W0 is the definition of the threshold. Since the weighted sum just needs to be a positive number, and 5 isn't negative, then it should still ~~fire~~ for H because the sum will just be even greater than 1 than before. The negative case doesn't change what it defines either; it just flips the co-orientation.

don't use the word fire here

2/2

2d: Not for the first case, but in the second case, the neuron will fire differently because now the sign will be flipped, so it'll be opposite, though it will still be defined the same way.

1/1

2e: No, because you're dividing by a constant for all weights, even though Wn could be something different from neuron to neuron, it's the same number in the denominator for every weight in any given neuron since it is its own Wn.

1/1

2f: No, because W0 is affected as well for H underscore, thus affecting the threshold. It could change the firing rule for H because if the weighted sum's result is positive or negative, that only won't change by being divided by a real number if that real number is positive, but Wn might be negative and in that case, all of the weights' signs will flip.

2/2

3a: For one dimension, a hyperplane is just a dot. The formula is the summation of M choose i, where i ranges from 0 to n. This sets the maximum upper-bound of regions that can exist in n space with m hyperplanes. So in this case, since n = 1, we would have m choose 0, always 1, plus m choose 1, which is m. So the result is at most m + 1 regions. So if you have 5 hyperplanes, or dots in this case, you'd have 5 + 1 regions.

.75/1

3b: Because you have a certain number of "degrees of freedom" to place your hyperplanes based off of the number of dimensions. As you have more dimensions, you can place more hyperplanes. The hyperplanes have to slice the dimensional space into a specific number of regions. So, if you have 0 dimensions, then you only have a single region, hyperplanes or otherwise. Adding another dimension, making it be n = 1, allows you to add many more regions. Overall, geometrically, you're choosing the number of hyperplanes m from among the possible dimensions, n, by definition, in order to get the number of regions. This is the definition of the binomial coefficient. • convoluted, misses basic explanation

2.75/3

3c: This is because if you expand the summation of the binomial coefficients, you'll get a polynomial that looks like this:

$$\frac{(m^3+5m+6)}{6} \text{ if } n=3$$

As m gets very large, the term that will grow fastest is the m cubed term, therefore, we can make a very accurate approximation if and only if m is large by considering only the highest power term, which will always be m raised to the nth power.

your intuition is there, write out the math

Extra credit: With sauer-shelah lemma, we know that the cardinality of every finite set family F is greater than the sum of n choose i, from i = 0 to k-1. With k being the number of dimensions and n being the number of items. We can also see this in O(n^k). All of this in our terms would be O(m^n) and |F| > the summation with i=0 to n-1 of m choose i. If the VC dimension of F is n, then we'd have that summation to n (not n-1) equaling O(m^n). This, of course, would expand to m choose 0 + m choose 1 + … + m choose n. If there is no hyperplane that divides n space into 2^(n+1) regions, then the upper bound must be 2^n. Considering that we know that the maximum number of regions with the constraint of the VC dimension of F being n in our terms is the aforementioned summation and that we know that we can't have that exactly, our maximum number of regions must be less than or equal to this formula.

+1.5 for attempt

- it sounds like you got it right, but it's hard to tell without seeing the math written
- great job!