BME118/218 Math Homework 1

Due April 19, 11:59PM

April 14, 2023

Problem 1. In Chapter 2 the text says "The main axioms of a vector space are that for any vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$, $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$ (associativity), $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$ (commutativity), and there exists a zero vector $\mathbf{0}$ such that for all vectors $\mathbf{v}$, $\mathbf{v}+\mathbf{0} = \mathbf{v}$. The rest of the axioms involve how the vectors are multiplied by scalars, and are exactly what you would expect." What are the rest of the axioms? Use letters a, b, c to denote scalars.

Inverse, Compatability, Identity
Distrib. w/ vect. addittion & multiplication

(answer for 1a)

wiki - vector space

What kind of a mathematical structure do the scalars need to form for them to be scalars of a vector space? Your choices are (1) a group, (2) a ring, (3) a field, or (4) the real numbers. Justify your choice.

no scalar division

a Field

wiki - vector space

(answer for 1b)

Note: you can use Wikipedia or Wolfram Alpha, or any other reference sources to answer this or any other question.

Problem 2. In the notation used in Chapter 3, let $\underline{\mathbf{x}} \stackrel{\text{def}}{=} (x_1, \ldots, x_n)^T$ represent an input vector to a neuron with n synapses and let $\mathbf{x} \stackrel{\text{def}}{=} (x_0, x_1, \ldots, x_n)^T$ represent that same input vector to that same neuron but with an extra coefficient $x_0 = 1$, i.e. x_0 is not a variable, but is always set to 1. Using the n-dimensional input vector $\underline{\mathbf{x}}$, suppose that neuron represents a hyperplane $\underline{H}$ in n-space defined by $\underline{H} \stackrel{\text{def}}{=} \{\underline{\mathbf{x}} = (x_1, \ldots, x_n)^T : \sum_{i=1}^{n} w_i x_i = -w_0\}$ where $w_1, \ldots, w_n$ represent synaptic strengths and w_0 is a quantity called the bias. The related quantity $T \stackrel{\text{def}}{=} -w_0$ is called the threshold. The neuron fires on inputs $\mathbf{x}$ where $x_0 = 1$ and $\sum_{i=1}^{n} w_i x_i > T$. Call this the *$n$-dimensional firing rule.*

Using the $n + 1$-dimensional input $\mathbf{x}$, that same neuron, with the same synaptic strengths and bias, represents a hyperplane H in $n+1$-space defined by $H \overset{\text{def}}{=} \{\mathbf{x} = (x_0, x_1, \dots, x_n)^T \text{ with } x_0 = 1 : \sum_{i=0}^{n} w_i x_i = 0\}$. Now the firing rule is that the neuron fires on inputs $\mathbf{x}$ where $\sum_{i=0}^{n} w_i x_i > 0$. Call this the $n + 1$-dimensional firing rule. From simple algebra, why do these two hyperplanes, $\underline{H}$ in n dimensions and H in $n + 1$ dimensions, represent the same firing rule for the neuron?

(answer for 2a) $\underline{H} \underset{\text{def}}{=} \{\dots\} = \{\dots : w_0 + \sum_{1}^{n} x_i w_i\} = \{\dots : (1) w_0 + \sum\dots\} = \{\dots : x_0 w_0 + \sum\dots\}$

$= \{\dots : \sum_{0}^{n} x_i w_i\} = H$

Assume $w_0 \neq 0$. What is the special property of H that $\underline{H}$ lacks?

H goes through origen

(answer for 2b)

Suppose $w_0, w_1, \dots, w_n$ are all multiplied by 5. Does either hyperplane change? Does either firing rule change? Suppose $w_0, w_1, \dots, w_n$ are all multiplied by -5. Does either hyperplane change?

(answer for 2c) no

Does either firing rule change? 5 unchanged

(answer for 2d) -5 changed

Suppose $w_n \neq 0$ and $w_0, w_1, \dots, w_n$ are all divided by w_n. Note that this reduces the number of free parameters needed to specify the hyperplane by 1, because $\frac{w_n}{w_n} = 1$. Does either hyperplane change under this "reduced parameter representation"?

(answer for 2e) no

Does either firing rule change?

(answer for 2f)

Problem 3. In the lecture we talked about the maximum number of regions that you can divide n-space into using m hyperplanes. These hyperplanes do not need to include the origin. Let's call this number $\Xi(n, m)$. Give an explicit formula for $\Xi(1, m)$ that you obtain by doing some examples and generalizing. Refer to a slide in the lecture for the definition of a hyperplane in 1 dimension.

(answer for 3a)

Geometrically, why is this formula right?

(answer for 3b)

In class we gave the formula $\Xi(3, m) = \binom{m}{0} + \binom{m}{1} + \binom{m}{2} + \binom{m}{3} \approx m^3$, and the general formula $\Xi(n, m) = \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n} \approx m^n$. Why is this approximation $\approx m^n$ ok for m a lot greater than n when all we care about is a so-called "big Oh" approximation? In other words, why is $\binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n} = O(m^n)$ for fixed n and variable m?

(answer for 3c)

Extra credit: Assume that you know that for any n, there is no set of $m \stackrel{\text{def}}{=} n+1$ hyperplanes that divide n space into $2^m \stackrel{\text{def}}{=} 2^{n+1}$ regions. Use this fact and the Sauer-Shelah Lemma as discussed in Wikipedia to prove that $\Xi(n, m) \leq \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n}$ for all m and n, not just $m = n + 1$.

Suppose $w_0, w_1, \ldots, w_n$ are all multiplied by 5. Does either hyperplane change?

Does either firing rule change? Suppose $w_0, w_1, \ldots, w_n$ are all multiplied by -5.

Does either hyperplane change? No

(answer for 2c) 1 check H

$$\sum 5 w_i x_i = -5 w_0$$
$$\Leftrightarrow 5 \sum w_i x_i = -5 w_0$$
$$\Leftrightarrow \sum w_i x_i = -w_0$$

2 check H

$$\sum 5 w_i x_i = 0$$
$$\Leftrightarrow 5 \sum w_i x_i = 0$$
$$\Leftrightarrow \sum w_i x_i = 0$$

according to definition of affine hyperplane these are equilent questions

Does either firing rule change?

(answer for 2d)

5 unchanged

−5 changed

Suppose $w_n \neq 0$ and $w_0, w_1, \ldots, w_n$ are all divided by w_n. Note that this reduces the number of free parameters needed to specify the hyperplane by 1, because $\frac{w_n}{w_n} = 1$. Does either hyperplane change under this "reduced parameter representation"?

(answer for 2e) No

Does either firing rule change?

(answer for 2f) Yes

if $w_n < 0$

1 check H

$$\sum_{1}^{n} \frac{w_i}{w_n} x_i = -w_0/w_n$$

$$\frac{1}{w_n} \sum_{1}^{n} w_i x_i = \frac{-w_0}{w_n}$$

2 check H

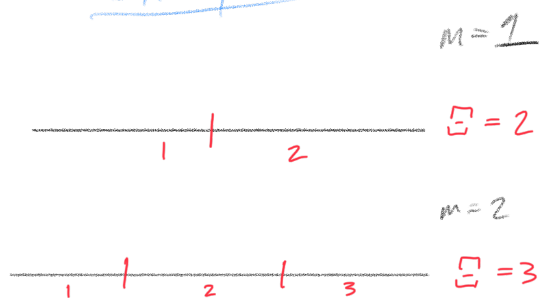
$$1/w_n \sum_{0}^{n} w_i x_i = 0$$

Problem 3. In the lecture we talked about the maximum number of regions that you can divide n-space into using m hyperplanes. These hyperplanes do not need to include the origin. Let's call this number $\Xi(n, m)$. Give an explicit formula for $\Xi(1, m)$ that you obtain by doing some examples and generalizing. Refer to a slide in the lecture for the definition of a hyperplane in 1 dimension.

(answer for 3a)

$$\boxed{\begin{aligned}\Xi(1,m) &= \binom{m}{0} + \binom{m}{1} \\ &= \frac{m!}{m!} + \frac{m!}{(1)(m-1)!} \\ &= 1 + m\end{aligned}}$$

examples:



3b

why does this make geometric sense?

- 1st hyperplane creates 2 regions
- each additional hyperplane adds region

In class we gave the formula $\Xi(3, m) = \binom{m}{0} + \binom{m}{1} + \binom{m}{2} + \binom{m}{3} \approx m^3$, and the general formula $\Xi(n, m) = \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n} \approx m^n$. Why is this approximation $\approx m^n$ ok for m a lot greater than n when all we care about is a so-called "big Oh" approximation? In other words, why is $\binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n} = O(m^n)$ for fixed n and variable m?

(answer for 3c)

- expanding left side of equation:

$$\Theta\left[\binom{m}{0} + \binom{m}{1} + \ldots + \binom{m}{n}\right]$$

$$= \Theta\left[\underbrace{1 + m + \ldots +}_{\text{this all becomes irrelavent in } \Theta} \frac{m(m-1)(m-n+1)}{n!}\right]$$

$$= \Theta\left[\frac{\overbrace{m(m-c)\ldots(m-c)}^{\text{multiplication happens } n \text{ times, constants } c \text{ become irrelavent}}}{c}\right]$$

$c \rightarrow$ constant becomes irrelavent for $n << m$

$$= \Theta[m \cdot m \cdot \ldots \cdot m]$$

$$= \Theta[m^n]$$

Suppose $n=2$

$$\Xi = \binom{m}{0} + \binom{m}{1} + \binom{m}{2}$$

$$= 1 + m + \frac{m!}{2!\,(m-2)!}$$

$$= 1 + m + \frac{m(m-1)}{2}$$

$\approx m^2$

Extra credit: Assume that you know that for any n, there is no set of $m \stackrel{\text{def}}{=} n+1$ hyperplanes that divide n space into $2^m \stackrel{\text{def}}{=} 2^{n+1}$ regions. Use this fact and the Sauer-Shelah Lemma as discussed in Wikipedia to prove that $\Xi(n, m) \leq \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n}$ for all m and n, not just $m = n + 1$.

replace the word "regions" with the word set

Definitions and statement [edit]

If $\mathcal{F} = \{S_1, S_2, \ldots\}$ is a family of sets, and T is another set, then T is said to be shattered by $\mathcal{F}$ if every subset of T (including the empty set and T itself) can be obtained as an intersection $T \cap S_i$ between T and a set in the family. The VC dimension of $\mathcal{F}$ is the largest cardinality of a set shattered by $\mathcal{F}$.

In terms of these definitions, the Sauer–Shelah lemma states that if $\mathcal{F}$ is a family of sets with m distinct elements such that

$$|\mathcal{F}| > \sum_{i=0}^{n-1} \binom{m}{i},$$

then $\mathcal{F}$ shatters a set of size n. Equivalently, if the VC dimension of $\mathcal{F}$ is n, then $\mathcal{F}$ can consist of at most

$$\sum_{i=0}^{n} \binom{m}{i} = O(m^n)$$

sets, as expressed using big O notation.

The bound of the lemma is tight: Let the family $\mathcal{F}$ be composed of all subsets of $\{1, 2, \ldots, m\}$ with size less than n. Then the size of $\mathcal{F}$ is exactly $\sum_{i=0}^{n-1} \binom{n}{i}$ but it does not shatter any set of size n.[7]