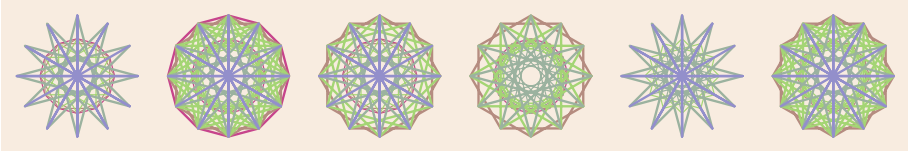
Review

CSC316 Machine Learning
Professor Leon Tabak

21 February 2022



1. We can multiply two matrices A and B if the number of columns in A matches the number of rows in B.

   The product AB will be a matrix with the same number of rows as A and the same number of columns as B.

   The number of rows in any matrix M^T is the same as the number of columns in M.

   What is the shape of $M^T M$?

2. The Normal Equation is a direct way of producing a linear regression. It calls for the computation of the inverse of a matrix. However, the fit() method of sklearn's LinearRegression does not compute an inverse. What does it do instead?

3. If A, B, and C are all matrices and $AB = C$, then how are the i^{th} row of A and the j^{th} column of B related to the element in the i^{th} and j^{th} column of C?

4. Here is the logistic function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

(a) Is the value of $f(1000)$ very close to 1.0, very close to 0.0, or equal to 0.5?

(b) Is the value of $f(-1000)$ very close to 1.0, very close to 0.0, or equal to 0.5?

(c) Is the value of $f(0)$ very close to 1.0, very close to 0.0, or equal to 0.5?

5. How does logistic regression differ from linear regression?

   - Think of when it might be appropriate to use each kind of model.
   - Think of how we compute a value using each kind of model.
   - Think of what kind of output each model produces. Or, put another way, how will we interpret the output in each case?

6. Some authors use the words "logistic function" and "sigmoid function" to mean the same thing. For others, a sigmoid function is a class of functions and the logistic function is just one example of a sigmoid function.

   (See, for example, articles on MathWorld and Wikipedia.)

   Another sigmoid function is the hyperbolic tangent:

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

(a) What value does $\tanh(x)$ approach if $x \gg 0.0$?

(b) What value does $\tanh(x)$ approach if $x \ll 0.0$?

(c) What is the value of $\tanh(0)$?

7. What do characteristics do the logistic function and the hyperbolic tangent have in common?

8. Our definition of entropy comes from Claude Shannon. Take a few minutes to read online about the life and work of Claude Shannon. What were his most important contributions to science, mathematics, or engineering?

9. To understand the definition of entropy, we need to know something about logarithms. Let's refresh our memories.

(a) What is the value of $\log_{10}(1000)$?

(b) What is the value of $\log_{2}(1024)$

(c) Does it seem reasonable that $\log_{10}(1024)$ will be close to 3?

(d) Does it seem reasonable that $\log_{10}(2)$ will be close to 1/3?
The logarithm base 10 of 2 is the power to which we must raise 10 to get 2.
Raising a number to the power 1/3 means computing its cube root.
Can you see that, since $2^3 = 8$ and 8 is close to 10, the cube root of 10 is probably a little more than 1/3?

(e) If...

$$\log_{10}(1024) \approx 3$$
$$\log_{10}(2) \approx \frac{1}{3}$$

What must be the approximate value of the one value divided by the other?

$$\frac{\log_{10}(1024)}{\log_{10}(2)} \approx \frac{3}{\frac{1}{3}}$$
$$\approx \ ?$$

(f) Try running this code in the Python IDLE shell:

```
1  import math
2
3  math.log10( 1024 ) / math.log10( 2 )
```

10. Let...

$$x = \log_2(a)$$
$$y = \log_2(b)$$

This means...

$$2^x = a$$
$$2^y = b$$
$$2^x \ 2^y = a \ b = 2^{x+y}$$

$$\log_2(ab) = x + y$$
$$= \log_2(a) + \log_2(b)$$

If we already know $\log_2(a)$ and $\log_2(b)$, how can we easily compute $\log_2(a/b)$?