LING572 HW7: Math needed for Neural Networks
Due: 11pm on Feb 22, 2022

Q1 (12 points): Let $f'(x)$ denote the derivative of a univariate function $f(x)$ w.r.t. the variable x.

(a) 2 pts: What does f'(x) intend to measure?
f'(x) intends to measure the rate of change of f(x) with respect to x.

(b) 2 pts: Let $h(x) = f(g(x))$. What is $h'(x)$ in terms of f'(x) and g'(x)?

$$h'(x) = g'(x) f'(g(x))$$

(c) 2 pts: Let $h(x) = f(x)g(x)$. What is $h'(x)$?

$$h'(x) = g'(x)f(x) + g(x)f'(x)$$

(d) 3 pts: Let $f(x) = a^x$, where $a > 0$. What is $f'(x)$?

$$f'(x) = a^x \ln(x)$$

(e) 3 pts: Let $f(x) = x^{10} - 2x^8 + \frac{4}{x^2} + 10$. What is $f'(x)$?

$$f'(x) = 10x^9 - 16x^7 - 8x^{-3}$$

Q2 (18 points): The logistic function is $f(x) = \frac{1}{1+e^{-x}}$. The tanh function is $g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

(a) 6 pts: Prove that $f'(x) = f(x)(1 - f(x))$.

Proof: Let $p(x) = 1+e^{-x}$. Then $f(x) = (p(x))^{-1}$. $p'(x) = -e^{-x}$

$$\text{lhs} = f'(x) = -(-e^{-x})(1+e^{-x})^{-2}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\text{rhs} = f(x)(1-f(x))$$

$$= \frac{1}{1+e^{-x}} \cdot \frac{\cancel{1}+e^{-x}\cancel{-1}}{1+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2} = \text{lhs}$$

□

(b) 6 pts: Prove that $g'(x) = 1 - g^2(x)$.

Proof: Let $p(x) = e^x - e^{-x}$, $q(x) = e^x + e^{-x}$. Then $g(x) = p(x)\, q^{-1}(x)$

$$\text{lhs} = g'(x) = p'(x)\, q^{-1}(x) + p(x)(-1)\, q^{-2}(x)\, q'(x)$$

$$= (e^x + e^{-x})(e^x + e^{-x})^{-1} + (e^x - e^{-x})(-1)(e^x + e^{-x})^{-2}(e^x - e^{-x})$$

$$= 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2}$$

$$= 1 - g^2(x)$$

□

(c) 6 pts: Prove that $g(x) = 2f(2x) - 1$

Proof: $2f(2x) - 1$

$$= 2 \cdot \frac{1}{1 + e^{-2x}} - 1$$

$$= \frac{(1 - e^{-2x})\, e^x}{(1 + e^{-2x})\, e^x}$$

$$= \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$= g(x)$$

□

Q3 (45 points): Let f be a multi-variate function, and let x be one of the variables in f. Let us denote the partial derivative of f with respect to x by f'_x or $\frac{df}{dx}$ or $\frac{\partial f}{\partial x}$. Please answer the following questions:

(b) 3 pts: What is the partial derivative f'_x trying to measure?

f'_x measures the rate of change of f with respect to only x while assuming all other variable(s) remain constant.

(c) 3 pts: How do you calculate the gradient of f at a point z?

Vector with each partial derivative at point z.

If z is represented as (x,y), the gradient at z would be $< f'_x, f'_y >$

(d) 5 pts: Suppose that $x = g(t)$ and $y = h(t)$ are differentiable functions of t and $z = f(x, y)$ is a differentiable function of x and y. How do you calculate $\frac{\partial z}{\partial t}$ using the chain rule of partial derivatives?

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t}$$

(e) 6 pts: Let $f(x, y) = x^3 + 3x^2y + y^3 + 2x$.

What is f'_x? What is f'_y?

↓

$$f'_x = 3x^2 + 6xy + 2$$

$$f'_y = 3x^2 + 3y^2$$

What is the gradient of $f(x, y)$ at point (1, 2)?

$$f'(1,2) = < 17, 15 >$$

(f) 3 pts: Let $z = \sum_{i=1}^{n} w_i x_i$. What is $\frac{\partial z}{\partial w_i}$?

x_i

(g) 5 pts: Let $f(z) = \frac{1}{1+e^{-z}}$ and $z = \sum_{i=1}^{n} w_i x_i$.
What is $\frac{\partial f}{\partial z}$?

$$\frac{\partial f}{\partial z} = f(z)(1-f(z)) \qquad (Q2.a)$$

What is $\frac{\partial f}{\partial w_i}$?

$$\frac{\partial f}{\partial w_i} = \frac{\partial f}{\partial z} \cdot \frac{\partial z}{\partial w_i} = f(z)(1-f(z))\, x_i$$

Hint: Use chain rule and your answers should contain $f(z)$.

(h) 5 pts: Let $E(z) = \frac{1}{2}(t - f(z))^2$, $f(z) = \frac{1}{1+e^{-z}}$ and $z = \sum_{i=1}^{n} w_i x_i$.

What is $\frac{\partial E}{\partial w_i}$? Hint: the answer should contain $f(z)$.

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial f} \cdot \frac{\partial f}{\partial z} \cdot \frac{\partial z}{\partial w_i}$$

$$= -(t - f(z)) f(z) (1 - f(z))\, x_i$$

Q4 (25 points): The softmax function: please read the short tutorial at
https://deepai.org/machine-learning-glossary-and-terms/softmax-layer
and answer the following questions:

(a) 2 pts: The softmax function is a function that takes the input x and produces the output y. What is the type of x? What is the type of y?
x is a vector of any real numbers. y is a vector with the same size as x and where the values add up to 1.

(b) 5 pts: In general which layer in neural network (NN) is the softmax function used and why?
Softmax function is generally used in the last layer in NN to convert the original output to a normaled probability distribution.

(c) 5 pts: What is the relationship between the softmax function and the sigmoid function?
Sigmoid function is a special case of softmax function. Sigmoid functions only take two input classes, so sigmoid and softmax work the same when there are two classes. When there are more than two classes, we should use softmax functions.

(d) 7 pts: What is the relationship between the softmax function and the argmax function? In NN, when do you use softmax and when do you use argmax?
Softmax function is a smoothed and differentiable alternative to the argmax function. Argmax functions return the greatest value as 1 and everything else as 0, while softmax function returns a probability distribution of the input values. In NN, softmax function is usually used in training and argmax is usually used in testing, since softmax preserves relevant information of the non-greatest value and so helps optimize a cost function and argmax easily returns a single predicted greatest value.

(e) 6 pts: If a vector x is [1, 2, 3, -1, -4, 0], what is softmax(x)? What is argmax(x)?
softmax(x):[0.08608, 0.23399, 0.63604, 0.01165, 0.00058, 0.03167]
argmax(x):[0,0,1,0,0,0]