

LING572 HW7: Math needed for Neural Networks

Due: 11pm on Feb 23, 2021

A few notes about this assignment:

- The answers to the questions should be pretty short. I've left some space for you to fill in the answers. I've also made the \LaTeX file available in case you want to add the answers to the latex file directly. In that case, you need to run `pdf2latex`, `latexmk`, or something like that to generate a pdf from the \LaTeX file.
- If you prefer to write formulas on paper (instead of typing them with \LaTeX or Word), it's ok. You just need to fill out the rest of the assignment, print out the file, insert formulas by hand, scan the paper, and then submit via Canvas.
- Since no programming is required, you only need to submit a single file. Please call it **readme.pdf**.
- The assignment has three parts:
 - Q1-Q2 are on the derivative of a univariate function (a function with a single variable), which should be covered in college-level calculus (a prerequisite of LING572).
 - Q3 is on the partial derivatives of a multivariate function. If you have not learned that topic before, you can look at the tutorials provided in Q3.
 - Q4 is on softmax, which might be new to you. I include the url of a short tutorial on the function.
- There are tons of textbooks and online tutorials that cover those topics. If the links provided in Q3-Q4 do not work for you or you are still confused after going over them, feel free to read any calculus textbook or search the Internet for more info.

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?

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

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

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

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

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 $f'(x) = f(x)(1 - f(x))$.

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

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

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:

(a) **15 free pts:** Refresh your memory about gradient, partial derivative, chain rule. Here are some readings on this. Free free to skip them if you already know the content. On the other hand, if you need more info or cannot access the videos as youtube is blocked in your country, just search for “partial derivatives”, “gradient”, and “chain rule with partial derivatives”. There should be tons of materials on those topics on the Internet.

- Khan Academy’s page on partial derivatives:
<https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivatives/a/introduction-to-partial-derivatives>
- Khan Academy’s page on the gradient:
<https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivatives/a/the-gradient>

- Chain rule with partial derivatives:
<https://tutorial.math.lamar.edu/classes/calci/chainrule.aspx>
- A 28-min tutorial on partial derivatives and the gradient:
https://www.youtube.com/watch?v=CnVes9TdnPo&ab_channel=TheOrganicChemistryTutor
- A 21-min video on Chain rule with partial derivatives:
https://www.youtube.com/watch?v=XipB_uExF0&ab_channel=TheOrganicChemistryTutor
- A one-hour tutorial on Partial derivatives:
https://www.youtube.com/watch?v=JAf_aSIJryg&ab_channel=TheOrganicChemistryTutor

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

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

(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?

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

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

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

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

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

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

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 $f(z)$.

Q4 (25 points): The softmax function: please read the short tutorial at <https://deeppai.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 ?

(b) **5 pts:** In general which layer in neural network (NN) is the softmax function used and why?

(c) **5 pts:** What is the relationship between the softmax function and the sigmoid function?

(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?

(e) **6 pts:** If a vector x is $[1, 2, 3, -1, -4, 0]$, what is $\text{softmax}(x)$? What is $\text{argmax}(x)$?

Submission: Submit the following to Canvas:

- Since HW7 has no coding part, you only need to submit your **readme.pdf** which includes answers to all the questions, plus anything you want TA to know. No need to submit anything else.