

## HW 5: Neural networks I and HW 6: Neural networks II

HW5 due Nov. 12 and HW6 due Nov. 19. Both are Thursdays and both at 11:59PM central time.

Help sessions for answering HW-related questions:

- HW5: Sunday, Nov. 8, 8-9PM, Wednesday, Nov. 11, 8-9PM, and office hours
- HW6: Sunday, Nov. 15, 8-9PM, Wednesday, Nov. 18, 8-9PM, and office hours

Pre-compiled PDF is [here](#). Command to compile on your own: `pandoc hw5_and_6.md -o hw5_and_6.pdf`

Please show intermediate steps for all computational problems below. Giving only the final result will result in zero point.

**For Problems 7 and above**, write steps in matrix form as long as you can to save your time. Do NOT detail sub-matrix steps – that’s a waste of time. You are encouraged to use computers to evaluate matrix operations rather than punching keys on a calculator. You are also encouraged to take advantage of the [MiniNN](#) library to do the computations for you.

**How to submit:** Just upload as PDF files to Canvas.

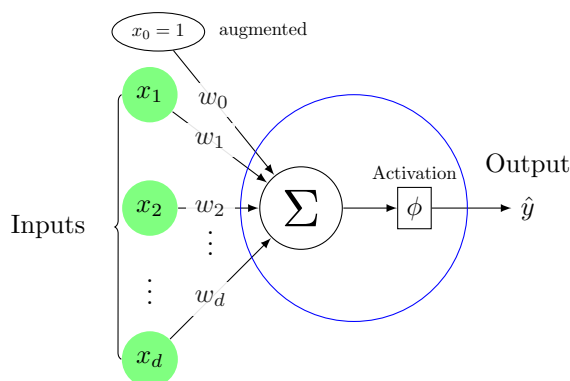
### HW 5: basic and single-neuron operations [10pts plus 4 bonus pts]

1. [1pt] What is the Hadamard product  $A \circ B$  between the following two matrixes?

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 0.5 & 0.1 & 0.3 \\ -1 & -20 & 1.5 \end{pmatrix}$$

2. [2pt] Continuing from Problem 1 above, what is the product  $AB^T$ ? And what is the product  $BA^T$ ?
3. [1pt] Continuing from Problems 1 and 2 above, is there a product  $AB$ ? Why?
4. [1pt] Continuing from Problems 1, 2, and 3, above, given  $f(x) = x + 1$ , what is the value of  $f(AB^T)$ ?
5. [Bonus, 2pt] In slides, to expand Eq. (2), we used negative logistic loss (also called cross entropy loss) as  $E$  and logistic activation function as  $\phi$ . What will be the result if we use squared error loss and linear activation function? Specifically, what if  $E = (\hat{y} - y)^2$  (assume just one sample) and  $\phi(\mathbf{w}^T \mathbf{x}) = \mathbf{w}^T \mathbf{x}$ ?
6. [2pt] Here is a diagram of a neuron.



Suppose  $d = 3$ . If the augmented input vector  $\mathbf{x} = [x_0, x_1, x_2, x_3] = [1, 0, 1, 0]^T$ , and the weight vector  $\mathbf{w} = [w_0, w_1, w_2, w_3] = [5, 4, 6, 1]^T$ , and the activation function  $\phi(x) = x^2$  (note that in function notation, the  $x$  in  $\phi(x)$  here can be any number or vector. not to be confused with the input vector  $\mathbf{x}$ ), what is the value of the prediction  $\hat{y}$ ?

Hint: Eq. (1)

7. [3pt] Continuing from problem 6 above, if the loss is defined as  $E = \hat{y} - y$ , what is the value of  $\partial E / \partial x_1$ ? And what is the value of  $\partial E / \partial w_1$ ?

Hint for second question: Eq. (2)

8. [Bonus, 2pt] What is the value of  $\frac{\partial E}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial E}{\partial x_0} \\ \frac{\partial E}{\partial x_1} \\ \vdots \end{pmatrix}$ ?

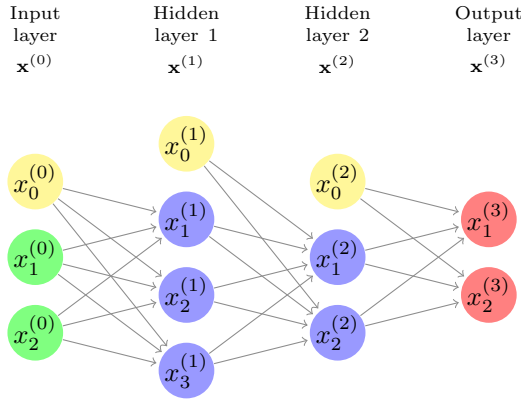
And what is the value of  $\frac{\partial E}{\partial \mathbf{w}} = \begin{pmatrix} \frac{\partial E}{\partial w_0} \\ \frac{\partial E}{\partial w_1} \\ \vdots \end{pmatrix}$ ?

Hint for second question: See the last equation on the same page with Eq. (2). But note that that  $E$  for that equation is neg log loss. Not the loss definition here.

## HW6: Operations on a neural network [10pts plus 5 bonus pts]

Hint: The slides “Recap:...” and “A grounded example...”.

9. [1pt] Here is a neural network.



Let  $\mathbb{W}^{(l)}$  be the transfer matrix from layer  $l$  to layer  $l + 1$ , for all  $l \in [0..2]$ .

What are the shapes (in terms of number of rows by number of columns, e.g.,  $5 \times 4$ ) for  $\mathbb{W}^{(0)}$ ,  $\mathbb{W}^{(1)}$ , and  $\mathbb{W}^{(2)}$  respectively?

10. [2pts] Continuing from Problem 9 above, if all weights in  $\mathbb{W}^{(0)}$  are 0.1, all weights in  $\mathbb{W}^{(1)}$  are 2, and all weights in  $\mathbb{W}^{(2)}$  are 1, what are the values of all activations  $\mathbf{x}^{(l)}$  for all  $l \in [1..3]$ ? Assume the input vector  $\mathbf{x}^{(0)} = [1, 1, 1]^T$ , the activation function be logistic function, and bias is 1  $x_0^{(l)} = 1, \forall l \in [0..2]$ . Express activations at each layer as a column vector.
11. [2.5pts] Continuing from Problems 9 and 10 above, if the target  $\mathbf{y}$  is  $[1, 0]^T$ , what are the values of  $\delta^{(l)}$  for all  $l \in \{2, 1\}$ ? Be sure to include  $\delta_0^{(l)}$  on the bias term if applicable. Suppose we use negative logistic (cross entropy) loss, and logistic activation function. Here  $\delta^{(3)} = \hat{\mathbf{y}} - \mathbf{y}$  is  $2 \times 1$  and the prediction  $\hat{\mathbf{y}} = \mathbf{x}^{(3)}$ .
12. [3pts] Continuing from Problems 9, 10, and 11 above, what are the values of  $\nabla^{(l)} = \frac{\partial E}{\partial \mathbb{W}^{(l)}}$  for all  $l \in [0..2]$ ?
13. [1.5pts] Finally, how should  $\mathbb{W}^{(l)}$  given in Problem 9 be updated to based on  $\nabla^{(l)}$  obtained in Problem 12, for all  $l \in [0..2]$ ?
14. [Bonus, 5pts] In [the demo for Unit 5 Regression](#), we used a neural network with tanh as the activation function for all neurons. The range of tanh is from -1 to 1, which means that the output from that neural network is limited between -1 and 1. But in that problem, the target or the prediction ranges from 0 to 4. How do you explain? Look into the source code of scikit-learn to find out.