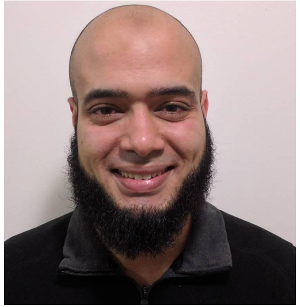
Machine Learning

Multi Classifier Homework Theoretical Qs

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Problem #1: Softmax Input Shifting

- Show that shifting the inputs of the softmax function by a constant C doesn't affect its results
  - Recall we shifted it with C = max(X) for numerical stability
- Tips
  - Write the equation with the shifting term C
  - Simplify in 1-2 lines

Problem #2: Softmax vs Sigmoid

- Recall for 2 classes classifier
  - Logistic regression ends with **a single logit** that we feed to sigmoid that refers to the positive class
  - However, softmax ends with **2 logits** that we feed to softmax
    - Z0 and Z1, where Z1 refers to the positive class
- With **simple 2-3 lines** of math, show the connection between softmax function and sigmoid when we have 2 classes
  - Specifically Softmax of Z1 versus a sigmoid function
- Tip: start with Softmax(z1) and simplify

Problem #3: Softmax Derivative

- Prove that the the partial derivative of the sigmoid function is based on 2 cases: one for the diagonal and one for non-diagonal
  - Below s(X) is the softmax function for vector X
    - s(X)_i: the ith output

$$\frac{\partial s(x)_i}{\partial x_j} = \begin{cases} s(x)_i(1 - s(x)_i) & \text{if } i = j \\ -s(x)_i s(x)_j & \text{if } i \neq j \end{cases}$$

Problem #3: Softmax Derivative

- Apply the quotient rule
  - See simple example
- Simplify with identifying y1 in mind
  - y1(1-y1) and   -y1y2

$$f(x) = \frac{u(x)}{v(x)}$$

$$= \frac{u'(x)v(x) - u(x)v'(x)}{v(x)^2}$$

$$y_1 = \frac{e^{x_1}}{e^{x_1} + e^{x_2} + e^{x_3} + e^{x_4}}$$

$$\frac{\partial y_1}{\partial x_1} = \frac{\partial\left(\frac{e^{x_1}}{e^{x_1}+e^{x_2}+e^{x_3}+e^{x_4}}\right)}{\partial x_1}$$

$$\frac{\partial y_1}{\partial x_2} = \frac{\partial\left(\frac{e^{x_1}}{e^{x_1}+e^{x_2}+e^{x_3}+e^{x_4}}\right)}{\partial x_2}$$

Problem #4: Softmax Derivative Implementation

- Write an **iterative code** to compute the softmax derivative from a given softmax probability output
- Then, think and write a **vectorized version**
  - Think in **simple way** how to rewrite the matrix
  - Tip: learn about [outer product]() and practice it numpy [np.outer]()

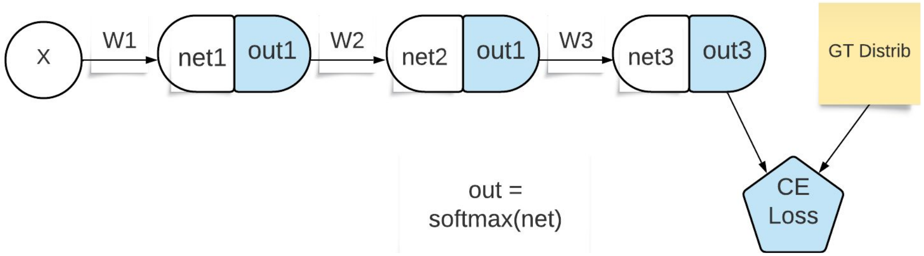
```
s = softmax(np.array([5, 7, 8]))
print(softmax_grad_iterative(s), '\n')
'''
[[ 0.03388568 -0.00911326 -0.02477242]
 [-0.00911326  0.19215805 -0.18304478]
 [-0.02477242 -0.18304478  0.2078172 ]]
'''
```

Problem #5: Softmax with Cross Entropy

- We learned in the lecture: ∂L / ∂net3 = out3 - gt3
- Derive that!
- Use these symbols
  - Let net3 = z
  - Let out3 = y' = softmax(z)
  - Let y = ground truth

$$L(y, \hat{y}) = -\sum_{i=1}^{K} y_i \log(\hat{y}_i)$$

$$\frac{\partial L}{\partial z_j} = \hat{y}_j - y_j$$



Problem #5: Softmax with Cross Entropy

- Start with the chain rule for differentiation
- You have 2 terms
  - The first term is direct based on log derivative
  - The second is the softmax derivative (2 cases
    - Put together the the 2 terms
    - Break the 2 cases of softmax derivative
    - Simplify
- Tip: Removing the sum symbol
  - For first term it is direct because of the derivative (i vs j)
  - For the 2nd term, use the fact that [The sum of y = 1]

$$\frac{\partial L}{\partial z_j} = \hat{y}_j - y_j$$

Problem #6: Kullback-Leibler (KL)-Divergence

- KL Divergence is a measure of how one probability distribution **diverges or is different** from a second, reference probability distribution (**dis**similarity)
  - Non-Negative / Not Symmetric
  - Unit: how different the two distributions (in bits for log2 and nats for loge)

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x} P(x) \log \frac{P(x)}{Q(x)}$$

- Prove that minimizing the cross entropy is equivalent to minimizing KL
  - Tip: In 2-3 lines decompose the above equation into some entropy and cross entropy
    - Recognize that one term is fixed during the training
    - Logically, formulate your final words

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."