MT Summer Term 2021 Ex10: Intro to NNs via Logistic Regression - Andrew Ng

1. In your own words, what is the difference between a loss function $\mathcal{L}$

$$\mathcal{L}(\hat{y}, y)$$

and a cost function $\mathfrak{J}(\mathbf{w}, b)$

$$\mathfrak{J}(\mathbf{w}, b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}\left(\hat{y}^{(i)}, y^{(i)}\right)$$

What are y and $\hat{y}$ and what does the index variable i in the cost function range over?

2. If your loss function $\mathcal{L}$ was the sum of squared errors

$$\sum_{j=1}^{n} \frac{1}{2}\left(y_j - \hat{y}_j\right)^2$$

for, say, an FFNN with n output neurons, what would the total cost function $\mathfrak{J}$ be? Explain the resulting cost function in your own words. Btw. why do have the $\frac{1}{2}$ in the sum of the squared errors?

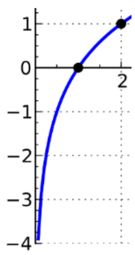
3. In your own words, what is the relationship between logistic regression and a single artificial neuron?

4. If you use logistic regression as a classifier, what is a popular loss function for it?

5. In your own words, explain how log/cross-entropy loss

$$\mathcal{L}(\hat{y}, y) = -(y \log \hat{y} + (1 - y)\log(1 - \hat{y}))$$

works. To do this, consider the case where the true class $y = 1$ and the case where $y = 0$. The following graph of the log may be useful:



6. Please complete the following:

$$\frac{d}{dx}\log_e x = \frac{d}{dx}\ln x = \dots$$

7. Please draw the computation graph of the following function

$$f(x, y, z) = 2(x^3 - 3(y + z))$$

using a suitable labelling of the sub-formulas in $f(x, y, z)$.

8. Please compute the forward pass on the computation graph for

$$f(x, y, z) = 2\left(x^3 - 3(y + z)\right)$$

where $x = 2, y = -1$ and $z = 3$.

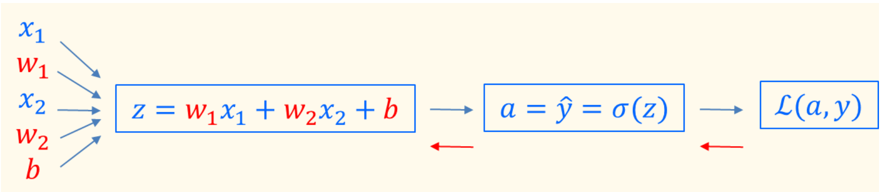
9. Please work out

$$\frac{\partial}{\partial x} 2\left(x^3 - 3(y+z)\right), \qquad \frac{\partial}{\partial y} 2\left(x^3 - 3(y+z)\right), \qquad \frac{\partial}{\partial z} 2\left(x^3 - 3(y+z)\right)$$

using the chain rule (and others, as required).

10. Please do the same as in (9), but this time use the computation graph and your sub-formula labelling from (7) above.

11. Given the following computation graph for logistic regression (a simple artificial neuron with sigmoid activation function):



please compute $\frac{\partial}{\partial a}\mathcal{L}(a,y)$, $\frac{\partial}{\partial z}\mathcal{L}(a,y)$, $\frac{\partial}{\partial w_2}\mathcal{L}(a,y)$ and $\frac{\partial}{\partial b}\mathcal{L}(a,y)$ using the computation graph.

12. Given the partial derivatives of the loss function with respect to weights in (11) above, in your own words please explain what the following does:

$$w_1 \coloneqq w_1 - \alpha \frac{\partial}{\partial w_1}\mathcal{L}(a,y)$$

$$w_2 \coloneqq w_2 - \alpha \frac{\partial}{\partial w_2}\mathcal{L}(a,y)$$

$$w_b \coloneqq w_b - \alpha \frac{\partial}{\partial w_b}\mathcal{L}(a,y)$$

13. Given the following definition of the softmax and sigmoid:

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{i=1}^{n} e^{x_i}}$$

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

and the following (row) vector of four real numbers:

$$\langle 2.41, -1.71, -0.16, 2.83 \rangle$$

please compute

i. elementwise sigmoid of the vector and
ii. a new vector where each element is the softmax of each of the elements in the original vector.

What does the softmax do to the original vector? What does the sigmoid do to each of the elements in the vector?

What do you think: is the softmax differentiable? I.e. can you backprop through a softmax?