Artificial Intelligence Course-2024
ECE- University of Tehran
Reference: Your Homework- Sunday (April 21, 2024, and April 28, 2024)

Part I

Xavier Glorot and Yoshua Bengio (2010) proposed an initialization scheme when training deep neural networks(https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf). Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun(2015) also proposed an initialization scheme when training neural networks(https://ieeexplore.ieee.org/document/7410480/authors#authors).

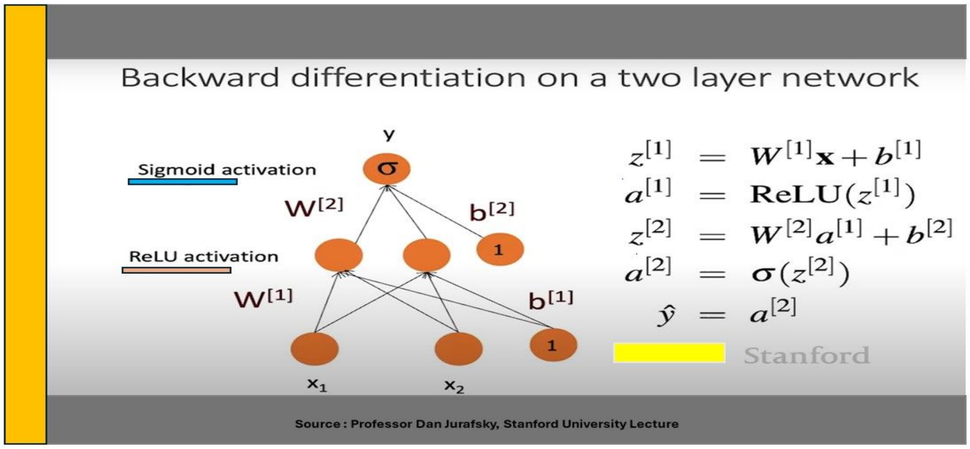
1. You should explain the mathematical intuition of both algorithms and compare them.
(3 students)
2. You should build two Deep Learning Models and compare your results. Use Python and TensorFlow with Google Colab. You should provide one notebook.
(3 students)

Part II

1. I mentioned in the lecture about the paper entitled [Towards Theoretically Understanding Why SGD Generalizes Better Than ADAM in Deep Learning]
https://proceedings.neurips.cc/paper/2020/file/f3f27a324736617f20abbf2ffd806f6d-Paper.pdf
by Pan Zhou, Jiashi Feng, Chao Ma, Caiming Xiong, Steven HOI, and Weinan E(2020). Is Stochastic Gradient Descent (SGD) preferred over Adaptive Moment Estimation (ADAM) in all Deep Learning applications? You must justify your answer by building two Deep Learning Models and employing SGD and ADAM. Use Python and TensorFlow with Google Colab. You should provide one notebook.
(3 students)

Part III

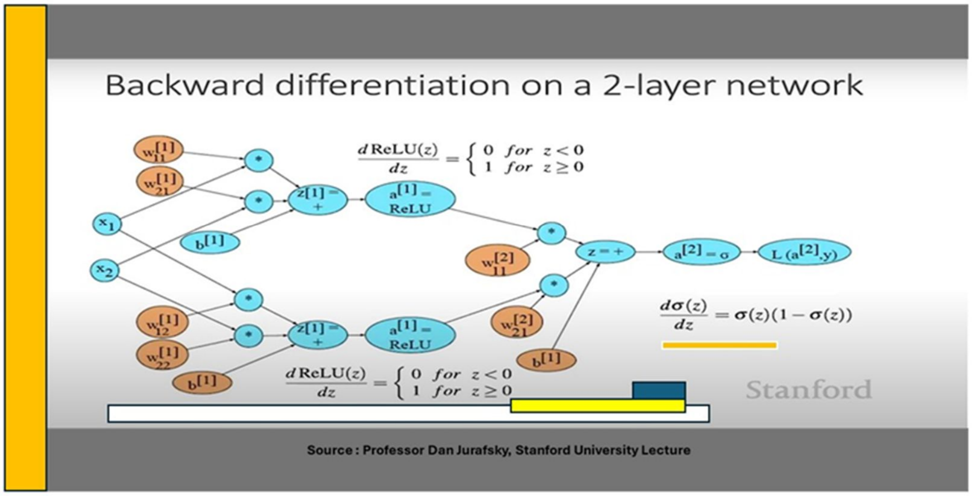
1. Show all the mathematical steps of the backward pass for the following two-layer neural network.
(2 students)



Backward differentiation on a two layer network

$$
\begin{aligned}
z^{[1]} &= W^{[1]}\mathbf{x} + b^{[1]} \\
a^{[1]} &= \mathrm{ReLU}(z^{[1]}) \\
z^{[2]} &= W^{[2]}a^{[1]} + b^{[2]} \\
a^{[2]} &= \sigma(z^{[2]}) \\
\hat{y} &= a^{[2]}
\end{aligned}
$$

Source : Professor Dan Jurafsky, Stanford University Lecture



Backward differentiation on a 2-layer network

$$\frac{d\,\text{ReLU}(z)}{dz} = \begin{cases} 0 & for \;\; z < 0 \\ 1 & for \;\; z \geq 0 \end{cases}$$

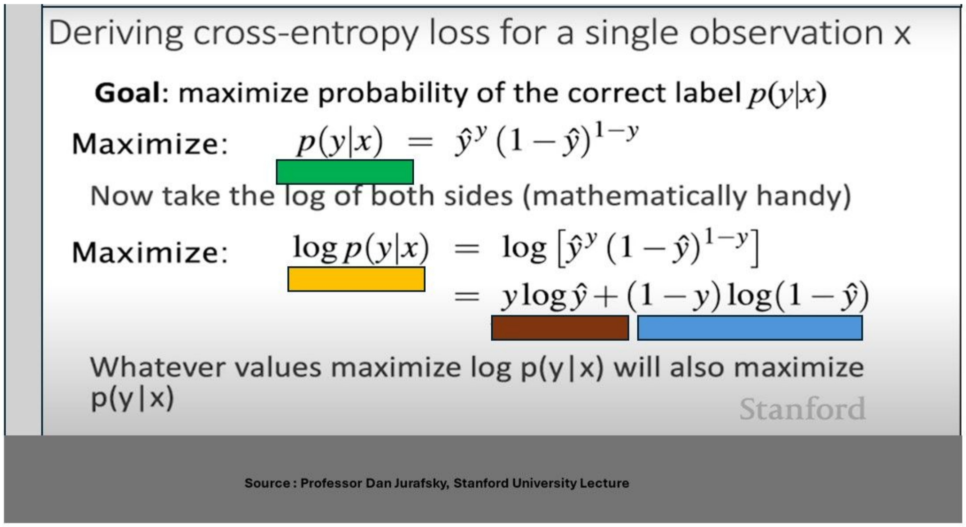
$$\frac{d\sigma(z)}{dz} = \sigma(z)(1 - \sigma(z))$$

$$\frac{d\,\text{ReLU}(z)}{dz} = \begin{cases} 0 & for \;\; z < 0 \\ 1 & for \;\; z \geq 0 \end{cases}$$

Source : Professor Dan Jurafsky, Stanford University Lecture

2. Can we say increasing the depth of our Neural Network model will always lead to optimum performance? You should answer this question with two examples. Use Python and TensorFlow with Google Colab. You should provide one notebook.
(2 students)

3. Explain the mathematical intuition for the cross-entropy loss of a single observation. Use the following.
(1 student)



Deriving cross-entropy loss for a single observation x

Goal: maximize probability of the correct label $p(y|x)$

Maximize: $$p(y|x) = \hat{y}^y (1 - \hat{y})^{1-y}$$

Now take the log of both sides (mathematically handy)

Maximize: $$\log p(y|x) = \log \left[\hat{y}^y (1 - \hat{y})^{1-y} \right] = y \log \hat{y} + (1 - y) \log (1 - \hat{y})$$

Whatever values maximize log p(y|x) will also maximize p(y|x)

Stanford

Source : Professor Dan Jurafsky, Stanford University Lecture

4. Explain the difference between Categorical cross-entropy and Sparse categorical cross-entropy for classification problems. You should provide two examples. Use Python and TensorFlow with Google Colab. You should provide one notebook.
(2 students)

5. What is the difference between Keras Sequential API and Keras Functional API? Use Python and TensorFlow with Google Colab. You should provide one notebook.
(2 students)

Part IV

1. Dimensionality Reduction is an important part of Machine Learning/Deep Learning. You should explain Principal Component Analysis, Linear Discriminant Analysis, and Isomap Embedding and compare these three algorithms for the dataset I have provided in the following Google Link.
Dataset https://drive.google.com/drive/folders/1j9PzDTaGyKr53z-J_kq6uC5yPdYsW_Qy
For this question, you should employ Deep Neural Networks. Note that the dataset is the same as the one you have used for your first assignment. Use Python and TensorFlow with Google Colab. You should provide one notebook.
(3 students)

Note

Dear Amir

Please divide the problems between the students as indicated in each question. I reckon we need two sessions for this homework. The first session will be on April 21, 2024, and the second session will be on April 28, 2024.

Cheers,