Handwritten Assignment-3
CSL2050 - Pattern Recognition and Machine Learning

NOTE:

1. This Handwritten Assignment contains ten problems. Total points for this assignment is 50. Please use A4 sheets and neatly solve all the problems in the proper order. Be precise, verbosity and messy writing will be penalized.
2. **(IMPORTANT)** Please write the following pledge in your handwriting with your signature on the first page of your assignment sheet (without this, the assignment will not be considered for evaluation):

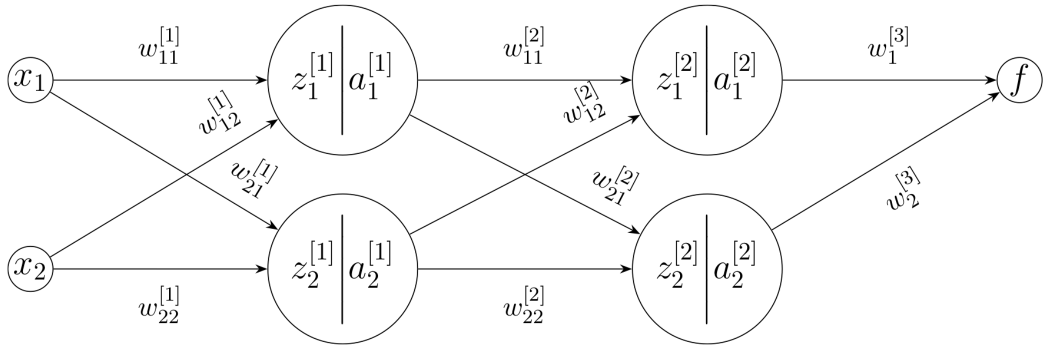
   Honesty Pledge:

   "I affirm that this assignment is solely my work. I have not used unauthorized assistance, engaged in plagiarism, or violated ethical standards. Further, all references used and any discussion with anyone have been appropriately cited. Any breach may lead to disciplinary actions as per the course academic honesty policy discussed in Lecture-1."
3. **Deadline:** April 30, 2024, 10:30 PM. Note: Please submit a **scanned handwritten sheet (soft copy)** in Google Classroom.
4. **Late Submission Policy:** Late submissions will not be considered for this assignment.

---

1. **(Backprop)** Consider a 3-layer network shown in Figure 1:
   Given that $f = w_1^{[3]}a_1^{[2]} + w_2^{[3]}a_2^{[2]}$. Compute the following derivatives: $\frac{\delta f}{\delta z_1^{[2]}}$, $\frac{\delta f}{\delta z_2^{[2]}}$, $\frac{\delta f}{\delta w_{11}^{[1]}}$, $\frac{\delta f}{\delta w_{22}^{[1]}}$ and $\frac{\delta f}{\delta w_{12}^{[1]}}$. Do the rough work separately and write down the final answer here. *(10 points)*

2. Draw a Computational Graph for the following function: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



$$Z^{[1]} = \begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad , \quad A^{[1]} = \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} = \begin{bmatrix} \sigma(z_1^{[1]}) \\ \sigma(z_2^{[1]}) \end{bmatrix}$$

$$Z^{[2]} = \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} = \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix} \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \quad , \quad A^{[2]} = \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix} = \begin{bmatrix} \sigma(z_1^{[2]}) \\ \sigma(z_2^{[2]}) \end{bmatrix}$$

Figure 1: Three Layer Network.

Now, assume $w = [w_0, w_1, w_2] = [1, 1, 1]$ and $x = [x_0, x_1] = [1, 1]$ and (a) perform the forward pass on the computation graph. (b) Compute local gradient times upstream gradients at every edge of the computation graph.

Hint: slide number 55-99 from `https://cs231n.stanford.edu/slides/2024/lecture_4.pdf` and video lecture: `https://www.youtube.com/watch?v=i94OvYb6noo&t=1787s`

3. Write down steps of k-means clustering algorithm. What are the limitations of the k-means clustering algorithm?

4. Given the following dataset in 1D space, which consists of 3 positive data points $\{-1, 0, 1\}$ and 3 negative data points $\{-3, -2, 2\}$.

(a) Find a feature map($R1 \rightarrow R2$), which will map the data in the ordinal 1-d input space (x) to a 2-d feature space (y1, y2) so that the data becomes linearly separable. Plot the dataset after mapping in 2D space.

(b) Write down the decision boundary $w_2y_2+w_1y_1+w_0$ given by hard-margin linear SVM in the feature space. Draw this decision boundary on your plot and mark the corresponding support vector(s).

(c) What is the equivalent decision boundary in the original input space? Draw this decision boundary over the original points.

5. This question concerns training Gaussian Mixture Models (GMM's). Throughout, we will assume there are two Gaussian components in the GMM. We will use μ_0, μ_1, σ_0 and σ_1 to define the means and variances of these two components and will use π_0 and $(1-\pi_0)$ to denote the mixture proportions of the two Gaussians (i.e., $p(x) = \pi_0\mathcal{N}(\mu_0,\sigma_0) + (1-\pi_0)\mathcal{N}(\mu_1,\sigma_1)$). We will also use to refer to the entire collection of parameters $(\mu_0, \mu_1, \sigma_0, \sigma_1, \pi_0)$ defining the mixture model $p(x)$.

(a) Prove that $p(x)$ is a valid PDF.

(b) Consider the training data set below and two clustering algorithms: K-Means and a Gaussian Mixture Model (GMM) trained using EM. Will these two clustering algorithms produce the same cluster centers (means) for this data set? In one-two sentence, explain why or why not.

(c) Consider following cluster centre initialization (μ_o as circle and μ_1 as star). In the next stage of expectation maximization, where will they shift and show the direction?

End of Paper