MatSoc, IITK

Architecting Intelligence

Assignment 1

Mentors: Basudev Mohapatra
Khushal Wadhwa
Mansi Ghodke

Due Date: December 15th, 2025

Total Marks: 40

1 Questions

Q1: Derivation of the Linear Regression Solution (8 marks)

You are given a dataset of n observations with the input features represented as a vector $x_i \in \mathbb{R}^d$, and scalar output/target value $y_i \in \mathbb{R}$.

The goal of linear regression is to find a parameter vector $\beta \in \mathbb{R}^{d+1}$ of the form:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id},$$

where β contains the intercept and all feature weights.

We define the **design matrix** $X \in \mathbb{R}^{n\times(d+1)}$ by stacking all input vectors and adding a bias column:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1d} \\ 1 & x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix}.$$

The first column consists of all 1's (for the intercept). Each subsequent column j contains the values of feature $x_{\cdot j}$ across all samples. We also define the prediction vector $\hat{y} = X\beta$

The **Least Squares Cost Function** is:

$$J(\beta) = \frac{1}{2}\|y - X\beta\|^2 = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

The factor $\frac{1}{2}$ is included for mathematical convenience. The **normal equation** is the analytic solution obtained by setting the gradient of $J(\beta)$ to zero.

Your Tasks:

(a) Explain, in your own words, what regression is doing and why minimizing squared error is a meaningful choice.

(b) Starting from the above cost function, derive the normal equation by computing the gradient with respect to β and setting it to zero.

(c) Discuss why direct inversion of $X^\top X$ in the normal equation may be problematic in high-dimensional or large-scale datasets, and explain why iterative methods (such as gradient descent) are often preferred.

Q2: Backpropagation in Neural Networks (8 marks)

Consider a feedforward neural network with:

- an input layer with 1 neuron, i.e input is a scalar $x \in \mathbb{R}$.
- one hidden layer with 2 neurons, each with sigmoid activation,
- one output layer neuron with sigmoid activation.

Forward pass for hidden neuron 1 is computed as:

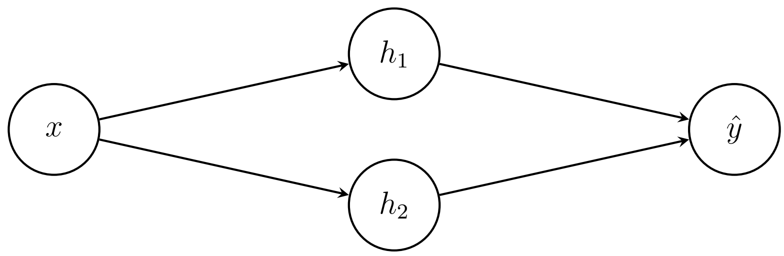
$$z_1 = w_1 x + b_1, \qquad a_1 = \sigma(z_1).$$

The output neuron receives inputs from both hidden neurons, but for this question we consider only the contribution of hidden neuron 1. The output neuron computes:

$$z_2 = w_2 a_1 + b_2, \qquad a_2 = \sigma(z_2) = \hat{y} \qquad \text{[Note: } \sigma(z) = \tfrac{1}{1+e^{-z}}\text{]}$$

The **binary cross-entropy (BCE)** loss for a single example (x, y) is:

$$\mathcal{L} = -\left(y \log(a_2) + (1-y)\log(1-a_2)\right).$$



Your tasks:

(a) Explain the core idea behind backpropagation and why the chain rule allows efficient computation of gradients in neural networks.

(b) Derive the gradients

$$\frac{\partial \mathcal{L}}{\partial w_2}, \qquad \frac{\partial \mathcal{L}}{\partial b_2}, \qquad \frac{\partial \mathcal{L}}{\partial w_1}, \qquad \frac{\partial \mathcal{L}}{\partial b_1}$$

for the network described above.
Hint: You might find it useful that $\dfrac{\partial \mathcal{L}}{\partial z_2} = a_2 - y$.

(c) How would you iteratively update the values of $w_1,\ w_2$ and $b_1,\ b_2$ in gradient descent? Briefly explain the role of the learning rate here.

Q3: ANN vs RNN vs LSTMs (5 marks)

Write a structured explanation comparing the three architectures. In particular, explain:

(a) how an ANN processes inputs compared to how an RNN processes sequences,

(b) why simple RNNs struggle with long-term dependencies,

(c) the role of gates in LSTMs and how they help preserve information,

(d) how LSTMs address the vanishing gradient problem,

(e) one example task that is best suited for each of ANN, RNN, and LSTM.

Q4: LSTMs in Natural Language Processing **(4 marks)**

(a) Give one example from natural language where correctly understanding a word or phrase depends on a "long-range dependency". Would a standard RNN struggle to model this dependency?

(b) Describe intuitively how the memory cell and the gating mechanism in an LSTM allow it to retain important information over long sequences. Describe a scenario in Machine Translation where the forget gate would need to be fully active (close to 0) to process the sequence correctly.

2 Coding

Q1: Character-Level Text Generation with LSTM **(15 marks)**

Objective: Implement a recurrent neural network architecture from scratch using PyTorch to generate text character-by-character.

Instructions: You are provided with the code for this part hosted on Google Colab.

- **Access the Notebook** by clicking here.
- **Create a Copy:** Upon opening the link, go to `File > Save a copy in Drive`. You cannot edit the original file; you must work on your own copy.
- **Enable GPU:** Before running any code, ensure you are using a GPU runtime to speed up training. Go to `Runtime > Change runtime type > T4 GPU`.

Required Tasks: The notebook provides a complete data pipeline and training loop. You are only required to complete the following three specific sections, marked clearly in the file:

- **Model Definition**: Locate the `CharLSTM` class. Implement the `__init__`, `forward`, and `init_hidden` methods using PyTorch.
- **Hyperparameters**: Adjust the configuration variables (Embedding Size, Hidden Size, Layers, Learning Rate) to achieve maximum model convergence.
- **Inference Configuration**: In the final "Probing Experiment" cell, play around with the `seed_text` and `gen_length` to generate unique text samples and see your model in action.

Each code cell is preceded by a detailed Markdown explanation. Read these to understand the logic before writing your code. While the data processing and training loops are fully functional, you may modify other parts of the notebook for further optimization if you wish to experiment.

Submission Guidelines: Download your completed notebook, with the trained model and all cells running completely, as an `.ipynb` file and submit this file along with the rest of the assignment.