

**WiDS ‘22 - ‘23 Final Documentation**

**22 - Diving Into ML**

**Soham Joshi and Anand Narasimhan**

|  |  |  |
| --- | --- | --- |
| **Team Member Name** | **Roll Number** | **Email-Id** |
| Ramswaroop | 203050117 | [203050117@iitb.ac.in](mailto:203050117@iitb.ac.in) |

**Introduction to Problem Statement**

|  |
| --- |
| The project's purpose is to establish a solid foundation in ML (its theory and basic applications) on which to build in the future. The project is divided into four weeks: week 0, week 1, week 2, and week 3. Each week included a new problem set.  Week0:  a) To calculate the time difference between iterative multiplication and matrix multiplication |

**Existing Resources**

|  |
| --- |
| **WEEK 0 -** <https://inside-machinelearning.com/en/quickly-upload-public-google-drive-files-on-notebook-and-colab/>  <https://www.analyticsvidhya.com/blog/2021/06/complete-guide-to-working-with-csv-files-in-python-with-pandas/>  <https://stackoverflow.com/questions/39662891/read-in-the-first-column-of-a-csv-in-python>  <https://colah.github.io/posts/2014-10-Visualizing-MNIST/>  <https://www.kaggle.com/code/yushg123/for-loops-vs-vectorized-who-wins-and-by-how-much/notebook>  **WEEK 1 -**  <https://explained.ai/matrix-calculus/>  [Calculus on Computational Graphs: Backpropagation -- colah's blog](https://colah.github.io/posts/2015-08-Backprop/)  [c++ - Is there any "standard" way to calculate the numerical gradient? - Stack Overflow](https://stackoverflow.com/questions/38854363/is-there-any-standard-way-to-calculate-the-numerical-gradient)  [Coding Neural Network — Forward Propagation and Backpropagtion | by Imad Dabbura | Towards Data Science](https://towardsdatascience.com/coding-neural-network-forward-propagation-and-backpropagtion-ccf8cf369f76)  [Of Gradients and Matrices. This artcle explains a line of code in… | by Paolo Perrotta | Medium](https://nusco.medium.com/of-gradients-and-matrices-1b19de65e5cd)  **WEEK 2 –**  [5.pdf (stanford.edu)](https://web.stanford.edu/~jurafsky/slp3/5.pdf)  [Logistic Regression as a Nonlinear Classifier – Data Exploration (xplordat.com)](https://xplordat.com/2019/03/13/logistic-regression-as-a-nonlinear-classifier/#:~:text=Logistic%20regression%20is%20known%20and%20used%20as%20a,to%20use%20logistic%20regression%20as%20a%20linear%20classifier.)  [numpy.random.Generator.normal — NumPy v1.25.dev0 Manual](https://numpy.org/devdocs/reference/random/generated/numpy.random.Generator.normal.html)  [Activation function - Wikipedia](https://en.wikipedia.org/wiki/Activation_function)  [Binary Classification – LearnDataSci](https://www.learndatasci.com/glossary/binary-classification/)  [matplotlib - plotting decision boundary of logistic regression - Stack Overflow](https://stackoverflow.com/questions/28256058/plotting-decision-boundary-of-logistic-regression)  [Logistic Regression and Decision Boundary | by Anuradha Wickramarachchi | Towards Data Science](https://towardsdatascience.com/logistic-regression-and-decision-boundary-eab6e00c1e8)  [Decision Boundary Visualization of Trained Logistic Regression (favtutor.com)](https://favtutor.com/blogs/decision-boundary-logistic-regression)  [F1 Score vs ROC AUC vs Accuracy vs PR AUC: Which Evaluation Metric Should You Choose? - neptune.ai](https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc) **WEEK 3 -** <https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi>  [Fundamentals of Neural Networks on Weights & Biases (wandb.ai)](https://wandb.ai/site/articles/fundamentals-of-neural-networks?https://wandb.ai/site/artifacts?utm_source=bing&utm_medium=cpc&utm_campaign=Conversions%3A+Marketing+Site+-+Non-Branded+-+Dynamic+Search&utm_content=Baseline+-+1st+Ad+India&msclkid=df0b01ce732e1f0975686214ba5bf2fe&utm_term=https%3A%2F%2Fwandb.ai%2Fsite)  [Linear Regression Model Neural Networks – Surfactants](https://www.surfactants.net/linear-regression-model-neural-networks/#:~:text=A%20linear%20regression%20model%20neural%20network%20is%20a,dependent%20variable%20is%20represented%20by%20a%20scalar%20value.)  [Choosing activation functions - Neural network training | Coursera](https://www.coursera.org/lecture/advanced-learning-algorithms/choosing-activation-functions-aWivF)  [PyTorch Activation Functions - ReLU, Leaky ReLU, Sigmoid, Tanh and Softmax - MLK - Machine Learning Knowledge](https://machinelearningknowledge.ai/pytorch-activation-functions-relu-leaky-relu-sigmoid-tanh-and-softmax/)  [Loss Functions in Neural Networks (theaidream.com)](https://www.theaidream.com/post/loss-functions-in-neural-networks)  [How to Train a Model with MNIST dataset | by Abdullah Furkan Özbek | Medium](https://medium.com/@afozbek_/how-to-train-a-model-with-mnist-dataset-d79f8123ba84)  [MNIST dataset using Deep Learning algorithm (ANN) | by Prateek Goyal | Medium](https://medium.com/@prtk13061992/mnist-dataset-using-deep-learning-algorithm-ann-c6f83aa594f5) |

**Proposed Solution**

|  |
| --- |
| Learning from available resources and understanding fundamental concepts try to solve each week problem. |

**Methodology & Progress (Mention the work done week-wise)**

|  |
| --- |
| **Week wise methodology and progress**  **Week0:**  a)   1. Measure the time it takes to perform a certain number of multiplications for iterative and matrix methods. 2. Compare the times obtained in above step and calculate the difference.   b)   1. Pre-processing - Convert the images to a standardized format and normalize the pixel values 2. Feature extraction - Extract important features from the images to be used as input to the model 3. Model training - Train a suitable machine learning model on the preprocessed data 4. Model evaluation - Evaluate the performance of the model on a separate test dataset 5. Model improvement - Fine-tune the model or try different models to improve its accuracy.   **Week1:**  **Problem 1)**   1. Analytic computation:    * If the function is defined in a symbolic form, use a library such as SymPy to symbolically differentiate the function with respect to each independent variable    * Calculate the gradient vector by combining the symbolic partial derivatives 2. Numeric computation:    * Create a loop to iterate over each independent variable    * Within the loop, perturb the variable by the step size, evaluate the function, and calculate the change in the function value    * Update the gradient approximation by dividing the change in the function value by the step size    * Repeat the process for each independent variable to obtain the gradient vector   **Problem 2)**   1. Forward Propagation:    * Pass the input data through the network, using matrix operations to compute the activations at each layer    * Apply activation functions at each layer to obtain the output of each layer    * Repeat the process until the final output layer is reached 2. Backward Propagation:    * Compute the error or loss between the predicted output and the true output    * Propagate the error backwards through the network, starting from the final output layer    * Compute the gradients of the loss with respect to the parameters of each layer    * Update the parameters to minimize the loss, using gradient-based optimization algorithms such as stochastic gradient descent   **Week2:**  Problem: Plotting the decision boundary   1. Prepare your data: Organize your data into input features and target labels, split into training and testing sets. 2. Train a classifier: Train a binary classifier such as a linear model, SVM, or decision tree on the training data. 3. Plot the decision boundary: To plot the decision boundary, you can use a mesh grid to generate a set of points that cover the space of the input features. You can then evaluate the classifier on these points to obtain a prediction for each point, which can be used to color the points based on their predicted label. Finally, you can plot the resulting-colored points on a graph to visualize the decision boundary. 4. Evaluate the classifier: Evaluate the performance of the classifier on the testing data to determine how well it generalizes to new data.   **Week3**   1. Load the MNIST data: Use a library such as TensorFlow or PyTorch to load the MNIST data set into memory. 2. Preprocess the data: Normalize the pixel values to be between 0 and 1, and one-hot encode the target labels. 3. Define the neural network architecture: Choose the number of hidden layers, the number of neurons in each layer, and the activation functions to use. 4. Compile the model: Define the loss function, optimizer, and metrics to use for training. 5. Train the model: Fit the model to the training data by specifying the number of epochs and batch size. Monitor the training loss and accuracy to ensure the model is converging. 6. Evaluate the model: Evaluate the model on the testing data to determine its generalization performance. 7. Fine-tune the model: Based on the evaluation results, adjust the architecture, optimizer settings, or other parameters to improve performance. 8. Use the model: Use the trained model to make predictions on new data. |

**Results**

|  |
| --- |
| https://github.com/dnyaneshwar9595/DK-Diving-into-ML.git |

**Learning Value**

|  |
| --- |
| **Got to know the fundamental concepts of ML as follows:**  ● Python (Numpy, Matplotlib, pandas, etc.) ● GitHub and google collaboratory ● Forward and backward propagation ● Computational graph ● Linear regression and binary classification ● Introductions Neural networks |

**Tech-stack Used**

|  |
| --- |
| Python (TensorFlow, PyTorch, scikit-learn) |

**Suggestions for others**

|  |
| --- |
| A really interesting and learning-oriented journey that provides hands-on Python programming as well as a basic introduction to Machine Learning and Neural Network ideas. The resources supplied were both instructive and practical. |

**Contribution by each Team Member**

|  |
| --- |
|  |

**References and Citations**

|  |
| --- |
| *1. https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.Perc eptron.html 2. https://www.w3schools.com/python/ 3. https://towardsdatascience.com/stochastic-gradient-descent-clearly-explain ed-53d239905d31 4. https://docs.python.org/3/library/timeit.html* |