

Assignment 1

Write your own backpropagation code and keep track of your experiments using wandb.ai

Mitesh Khapra

▼ Instructions

- The goal of this assignment is twofold: (i) implement and use gradient descent (and its variants) with backpropagation for a classification task (ii) get familiar with wandb which is a cool tool for running and keeping track of a large number of experiments
- We strongly recommend that you work on this assignment in a team of size 2. Both the members of the team are expected to work together (in a subsequent viva both members will be expected to answer questions, explain the code, etc).
- Collaborations and discussions with other groups are strictly prohibited.
- You must use Python (numpy and pandas) for your implementation.
- You cannot use the following packages from keras, pytorch, tensorflow: optimizers, layers
- If you are using any packages from keras, pytorch, tensorflow then post on moodle first to check with the instructor.
- You can run the code in a jupyter notebook on colab by enabling GPUs.
- You have to generate the report in the same format as shown below using wandb.ai. You can start by cloning this report using the clone option above. Most of the plots that we have asked for below can be (automatically) generated using the apis provided by wandb.ai. You will upload a link to this report on gradescope.

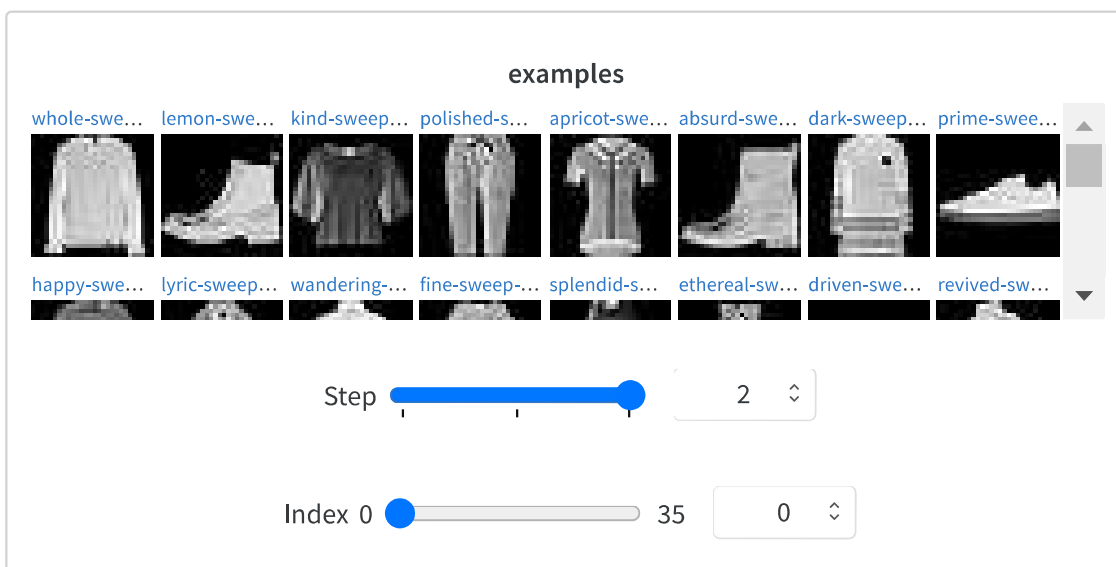
- You also need to provide a link to your github code as shown below. Follow good software engineering practices and set up a github repo for the project on Day 1. Please do not write all code on your local machine and push everything to github on the last day. The commits in github should reflect how the code has evolved during the course of the assignment.
- You have to check moodle regularly for updates regarding the assignment.

▼ Problem Statement

In this assignment you need to implement a feedforward neural network and write the backpropagation code for training the network. We strongly recommend using numpy for all matrix/vector operations. You are not allowed to use any automatic differentiation packages. This network will be trained and tested using the Fashion-MNIST dataset. Specifically, given an input image (28 x 28 = 784 pixels) from the Fashion-MNIST dataset, the network will be trained to classify the image into 1 of 10 classes.

▼ Question 1 (2 Marks)

Download the fashion-MNIST dataset and plot 1 sample image for each class as shown in the grid below. Use "from keras.datasets import fashion_mnist" for getting the fashion mnist dataset.



▼ Question 2 (10 Marks)

Implement a feedforward neural network which takes images from the fashion-mnist data as input and outputs a probability distribution over the 10 classes.

Your code should be flexible so that it is easy to change the number of hidden layers and the number of neurons in each hidden layer.

We will check the code for implementation and ease of use.

▼ Question 3 (18 Marks)

Implement the backpropagation algorithm with support for the following optimisation functions

- sgd
- momentum based gradient descent
- nesterov accelerated gradient descent
- rmsprop
- adam
- nadam

(12 marks for the backpropagation framework and 2 marks for each of the optimisation algorithms above)

We will check the code for implementation and ease of use (e.g., how easy it is to add a new optimisation algorithm such as Eve). Note that the code should be flexible enough to work with different batch sizes.

▼ Question 4 (10 Marks)

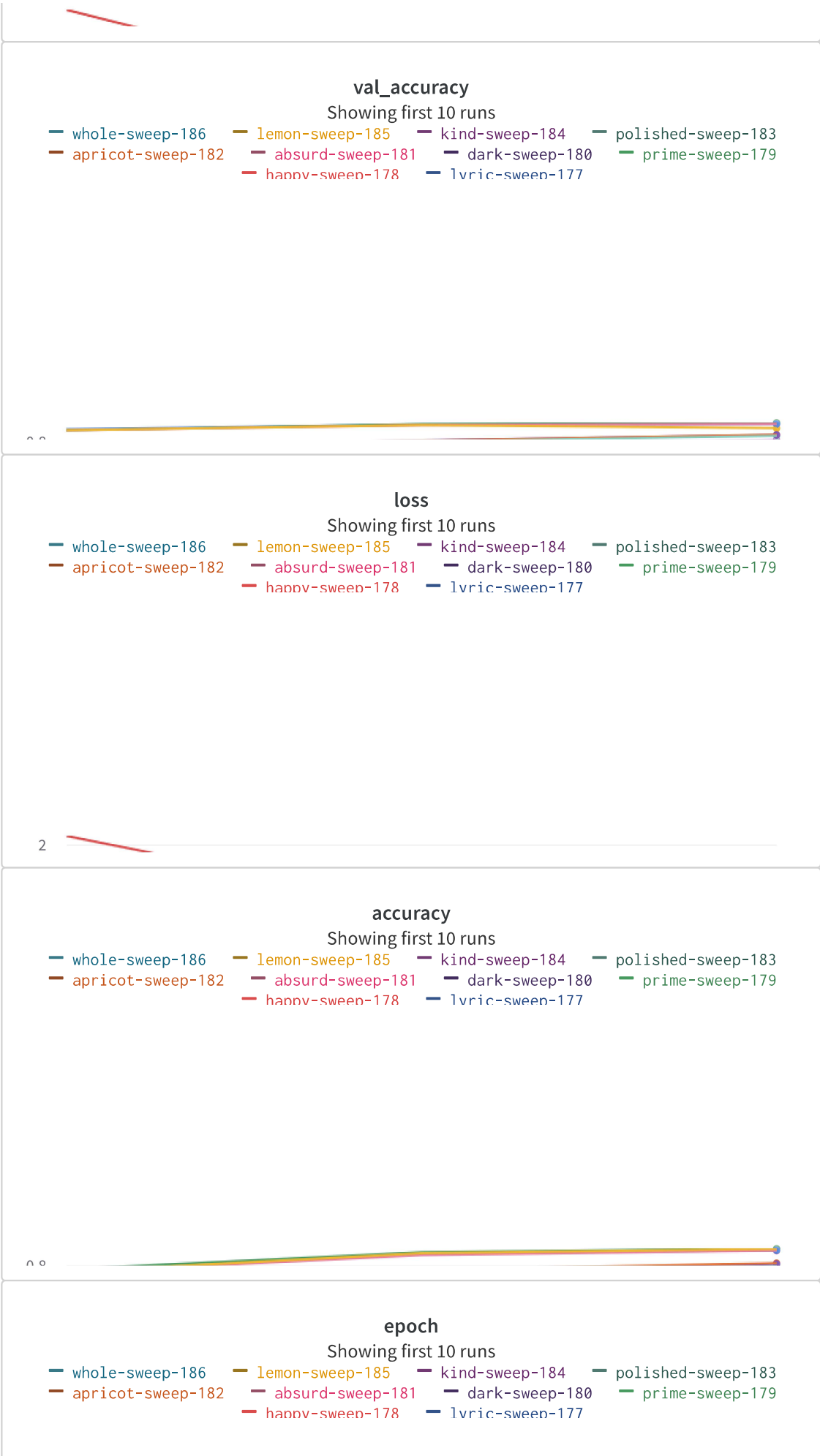
Use the sweep functionality provided by wandb to find the best values for the hyperparameters listed below. Use the standard

train/test split of fashion_mnist (use (X_train, y_train), (X_test, y_test) = fashion_mnist.load_data()). Keep 10% of the training data aside as validation data for this hyperparameter search. Here are some suggestions for different values to try for hyperparameters. As you can quickly see that this leads to an exponential number of combinations. You will have to think about strategies to do this hyperparameter search efficiently. Check out the options provided by wandb.sweep and write down what strategy you chose and why.

- number of epochs: 5, 10
- number of hidden layers: 3, 4, 5
- size of every hidden layer: 32, 64, 128
- weight decay (L2 regularisation): 0, 0.0005, 0.5
- learning rate: 1e-3, 1 e-4
- optimizer: sgd, momentum, nesterov, rmsprop, adam, nadam
- batch size: 16, 32, 64
- weight initialisation: random, Xavier
- activation functions: sigmoid, tanh, ReLU

wandb will automatically generate the following plots. Paste these plots below using the "Add Panel to Report" feature. Make sure you use meaningful names for each sweep (e.g. hl_3_bs_16_ac_tanh to indicate that there were 3 hidden layers, batch size was 16 and activation function was ReLU) instead of using the default names (whole-sweep, kind-sweep) given by wandb.

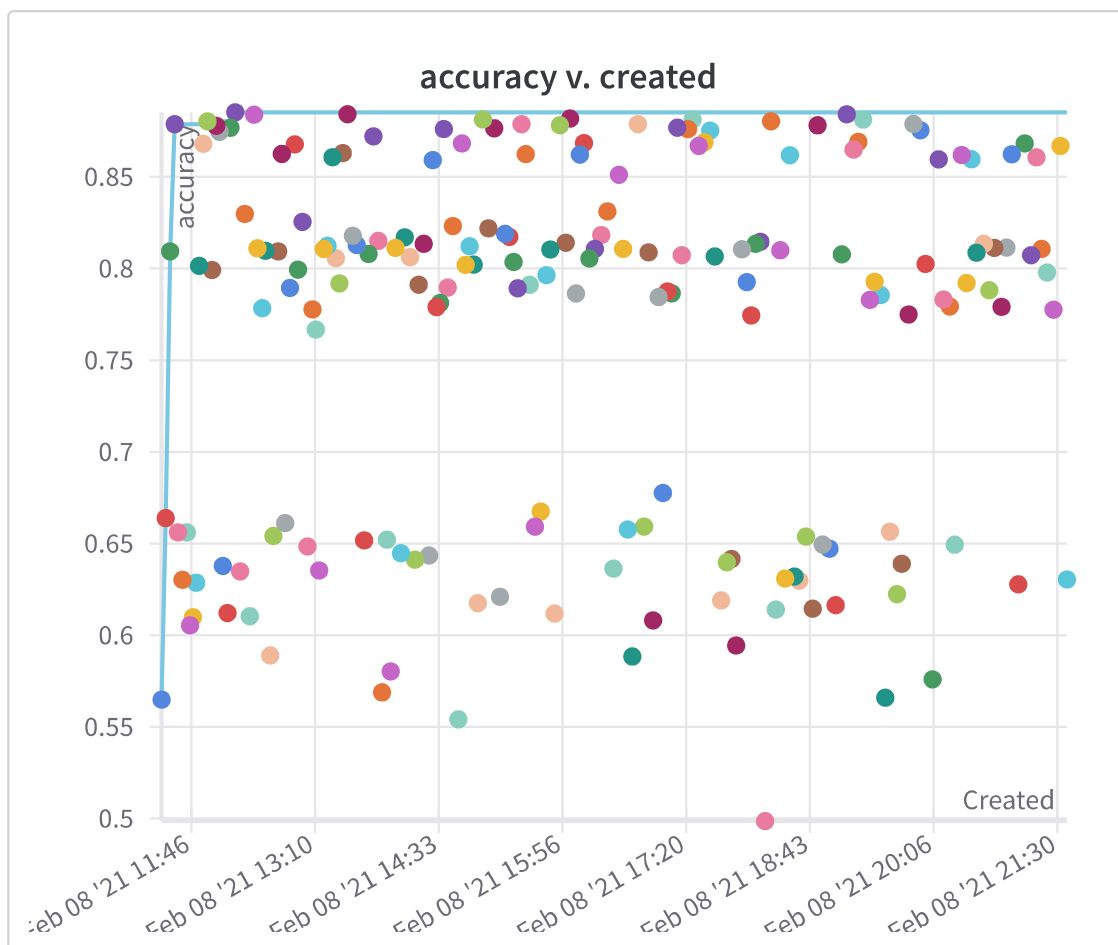




▼ Question 5 (5 marks)

We would like to see the best accuracy on the validation set across all the models that you train.

wandb automatically generates this plot which summarises the test accuracy of all the models that you tested. Please paste this plot below using the "Add Panel to Report" feature



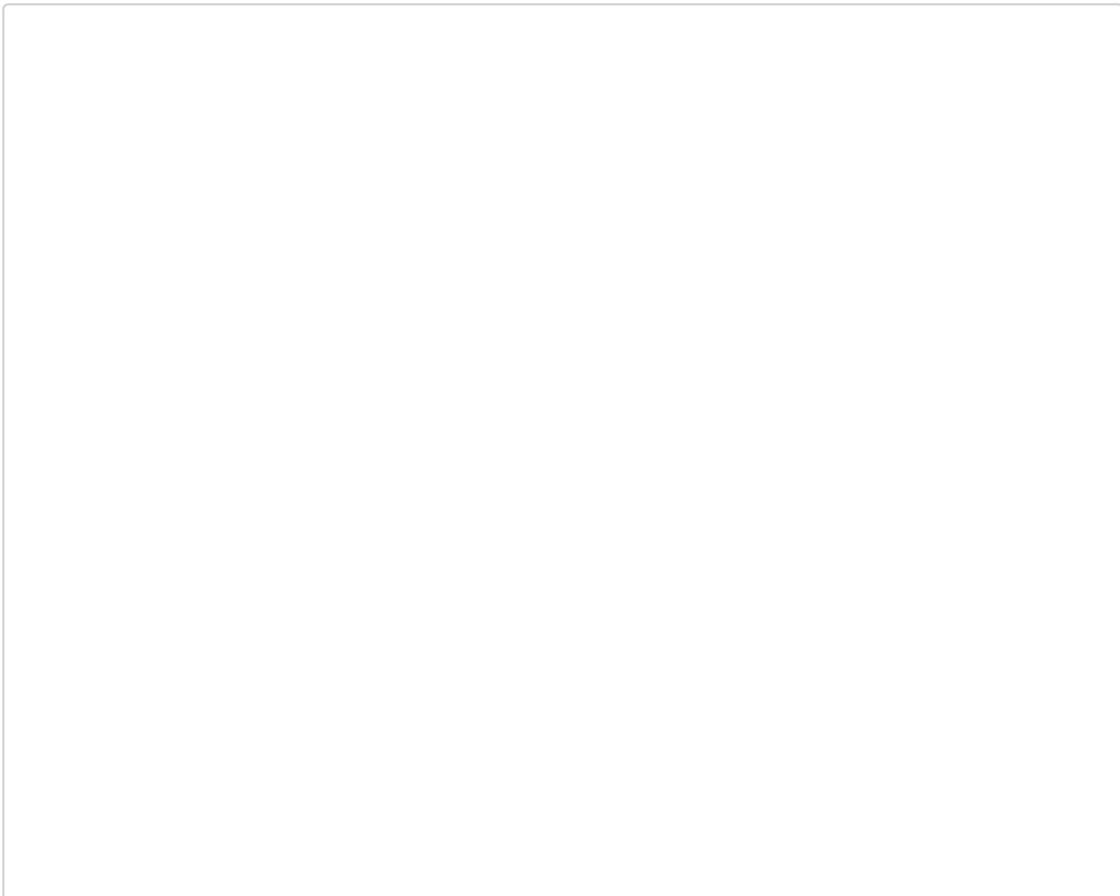
▼ Question 6 (20 Marks)

Based on the different experiments that you have run we want you to make some inferences about which configurations worked and which did not.

Here again, wandb automatically generates a "Parallel co-ordinates plot" and a "correlation summary" as shown below. Learn about a "Parallel co-ordinates plot" and how to read it.

By looking at the plots that you get, write down some interesting observations (simple bullet points but should be insightful). You can also refer to the plot in Question 5 while writing these insights. For example, in the above sample plot there are many configurations which give less than 65% accuracy. I would like to zoom into those and see what is happening.

I would also like to see a recommendation for what configuration to use to get close to 95% accuracy.

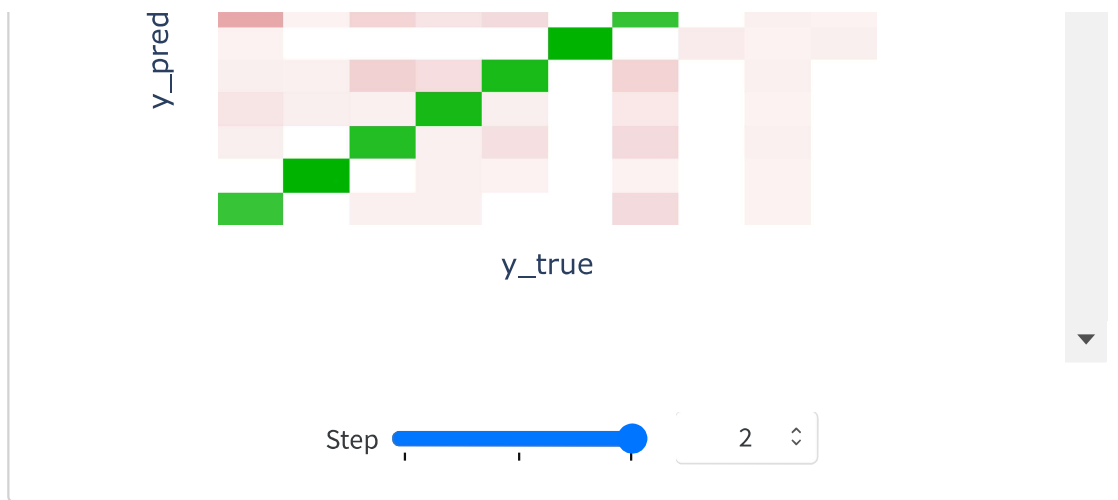


Parameter importance with respect to val_accuracy		
Q Search	Parameters	1-10 of 10 < >
Config parameter	Importance ⓘ ↓	Correlation
learning_rate		
optimizer.value_sgd		
optimizer.value_adam		
layer_2_size		
Runtime		
weight_decay		
layer_1_size		

▼ Question 7 (10 Marks)

For the best model identified above, report the accuracy on the test set of fashion_mnist and plot the confusion matrix as shown below. More marks for creativity (less marks for producing the plot shown below as it is)





▼ Question 8 (5 Marks)

In all the models above you would have used cross entropy loss. Now compare the cross entropy loss with the squared error loss. I would again like to see some automatically generated plots or your own plots to convince me whether one is better than the other.

▼ Question 9 (10 Marks)

Paste a link to your github code for this assignment

Example: https://github.com/<user-id>/cs6910_assignment1;

- We will check for coding style, clarity in using functions and a README file with clear instructions on training and evaluating the model (the 10 marks will be based on this)
- We will also run a plagiarism check to ensure that the code is not copied (0 marks in the assignment if we find that the code is plagiarised)
- We will check the number of commits made by the two team members and then give marks accordingly. For example, if we see 70% of the commits were made by one team member then that member will get more marks in the assignment (**note that this contribution will decide the marks split for the entire assignment and not just this question**).

- We will also check if the training and test data has been split properly and randomly. You will get 0 marks on the assignment if we find any cheating (e.g., adding test data to training data) to get higher accuracy

▼ Question 10 (10 Marks)

Based on your learnings above, give me 3 recommendations for what would work for the MNIST dataset (not Fashion-MNIST). Just to be clear, I am asking you to take your learnings based on extensive experimentation with one dataset and see if these learnings help on another dataset. If I give you a budget of running only 3 hyperparameter configurations as opposed to the large number of experiments you have run above then which 3 would you use and why. Report the accuracies that you obtain using these 3 configurations.

▼ Self Declaration

List down the contributions of the two team members:

For example,

CS20Mzzzz: (70% contribution)

- implementing sgd, nesterov, ...
- plotting the confusion matrix
- setting up the sweep in wandb
- analysing the parallel co-ordinates plot and writing inferences
- ...
- ...

CS20Myyyy: (30% contribution)

- implementing sgd, nesterov, ...
- setting up the sweep in wandb

- ...
- ...

We, Name_XXX and Name_ZZZ, swear on our honour that the above declaration is correct.

Note: Your marks in the assignment will be in proportion to the above declaration. Honesty will be rewarded (Help is always given in CS6910 to those who deserve it!).

This is an opportunity for you to come clean. If one of the team members has not contributed then it should come out clearly from the above declaration. There will be a viva after the submission. If your performance in the viva does not match the declaration above then both the team members will be penalised (50% of the marks earned in the assignment).

Created with  on Weights & Biases.

[https://wandb.ai/miteshk/assignments/reports/Assignment-1--VmlldzozNjk4NDE?
accessToken=r7ndsh8lf4wlxyjln7phvvfb8ftvc0n4lyn4tiowdg06hhzpzfzki4jrm28wqh44](https://wandb.ai/miteshk/assignments/reports/Assignment-1--VmlldzozNjk4NDE?accessToken=r7ndsh8lf4wlxyjln7phvvfb8ftvc0n4lyn4tiowdg06hhzpzfzki4jrm28wqh44)