

REPORT ON ASSIGNMENT 2 (MACHINE LEARNING LAB)

❖ PROCEDURES

★ CLASSIFICATION USING SINGLE-LAYER PERCEPTRON (SLP)

❑ ON BREAST-CANCER DATASET

- The dataset, available in scikit-learn was loaded and split into training and test sets in the ratio 7:3, with random shuffling.
- As of the model, the Perceptron model was loaded from scikit-learn and the model was trained on the data. Training was done for 1000 iterations. Predictions were made on the test data, and the accuracy was calculated by counting the number of correct predictions.
- The accuracy has been compared with a multi-layer perceptron having a single hidden layer too.

❑ ON HOUSE-VOTES 84 DATASET

- Same procedure as mentioned above was followed for this dataset also, and the accuracy obtained was compared against that obtained from multi-layer perceptron model with a single hidden layer.

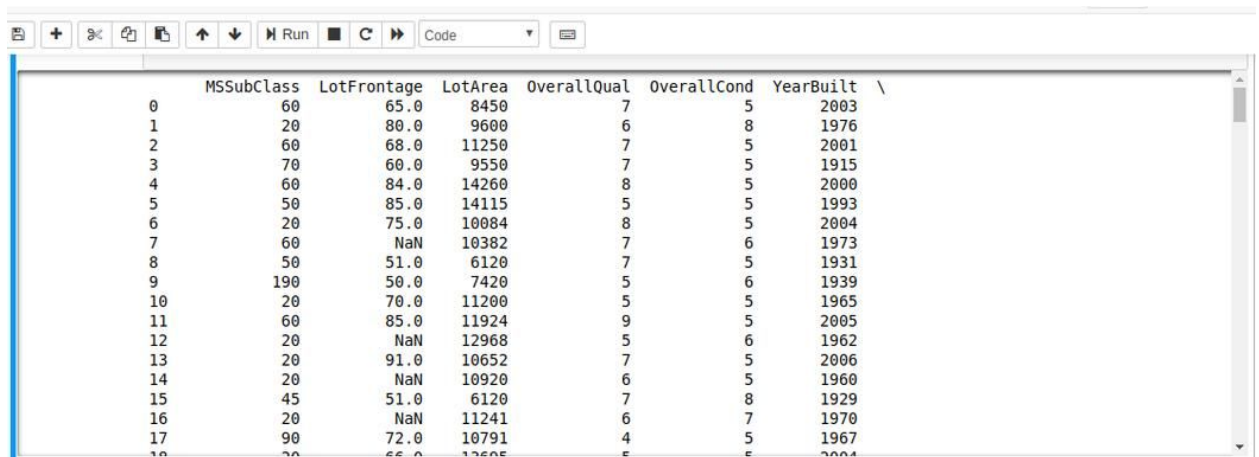
★ CLASSIFICATION USING MULTI-LAYER PERCEPTRON (MLP) ON BREAST-CANCER AND HOUSE-VOTES 84 DATASETS

- For both the datasets, the same procedure was followed as in SLP, except that this time, MLPs were used as the training models. First, for both the datasets, MLPs with one hidden layer were trained and their performance was compared to the SLPs trained previously.
- During training, two different optimizers were used and the results obtained - “Limited-Memory BFGS (LBFGS)” (as it works well with small datasets) and “Stochastic Gradient Descent (SGD). Training was done for 1000 iterations.

- The logistic loss plots and the respective accuracies were also obtained.
- Afterwards, for the house-votes 84 dataset, MLPs with one, two and three hidden layers were trained sequentially, and their performances were compared against each other. In this case too, both LBFGS and SGD were tried as optimizers.

★ REGRESSION ON HOUSE-PRICES DATASET USING MULTI-LAYER PERCEPTRON (MLP)

- The dataset was downloaded from the given link and the categorical attributes were manually converted to one-hot-encoded features - for each of the unique values of each categorical feature, a new column was created, and the original columns dropped; then a 1 was placed in the rows in which a particular value originally existed, and the rest of the rows for that column were set to 0. In this manner, the original dataset was preprocessed.



The screenshot shows a Jupyter Notebook interface with a table of house data. The table has 8 columns: MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, and YearBuilt. The rows are indexed from 0 to 17. The data is as follows:

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt
0	60	65.0	8450	7	5	2003
1	20	80.0	9600	6	8	1976
2	60	68.0	11250	7	5	2001
3	70	60.0	9550	7	5	1915
4	60	84.0	14260	8	5	2000
5	50	85.0	14115	5	5	1993
6	20	75.0	10084	8	5	2004
7	60	NaN	10382	7	6	1973
8	50	51.0	6120	7	5	1931
9	190	50.0	7420	5	6	1939
10	20	70.0	11200	5	5	1965
11	60	85.0	11924	9	5	2005
12	20	NaN	12968	5	6	1962
13	20	91.0	10652	7	5	2006
14	20	NaN	10920	6	5	1960
15	45	51.0	6120	7	8	1929
16	20	NaN	11241	6	7	1970
17	90	72.0	10791	4	5	1967

- 90% of train_data (as downloaded) was used as training set and 10% as test set.
- MLPs for one and two hidden layers were applied on the training set and predictions for house prices were done on the test set (again using

both LBFGS and SGD as optimizers). Finally the results were obtained.

1457	0.0	0.0	0.0	0.0	1.0
1458	0.0	0.0	0.0	0.0	1.0
1459	0.0	0.0	0.0	0.0	1.0
	SaleCondition_1	SaleCondition_2	SaleCondition_3	SaleCondition_4	\
0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	
2	0.0	0.0	0.0	0.0	
3	1.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	
5	0.0	0.0	0.0	0.0	
6	0.0	0.0	0.0	0.0	
7	0.0	0.0	0.0	0.0	
8	1.0	0.0	0.0	0.0	
9	0.0	0.0	0.0	0.0	
10	0.0	0.0	0.0	0.0	
11	0.0	1.0	0.0	0.0	
12	0.0	0.0	0.0	0.0	
13	0.0	1.0	0.0	0.0	
14	0.0	0.0	0.0	0.0	

- The mean and standard deviation for the error in prediction with respect to test set were calculated and the corresponding graphs were plotted.

❖ RESULTS

TABLE 1: ACCURACY COMPARISONS (IN %)

DATASET	SLP	MLP (1 HIDDEN LAYER)	MLP (2 HIDDEN LAYERS)	MLP (3 HIDDEN LAYERS)
Breast Cancer	93.57	98.24	-	-
House-Votes 84	95.42	93.9 (sgd) 96.18 (lbfgs)	64.88 (sgd) 94.65 (lbfgs)	64.88 (sgd) 96.94 (lbfgs)

**TABLE 2: STATISTICAL MEASURES FOR ERROR IN PREDICTION
(USING MLP FOR REGRESSION ON HOUSE-PRICES DATASET)**

MEASURE	MLP (1 HIDDEN LAYER)	MLP (2 HIDDEN LAYERS)
MEAN	5506.38	1515.04
STANDARD DEVIATION	38080.33	35199.19

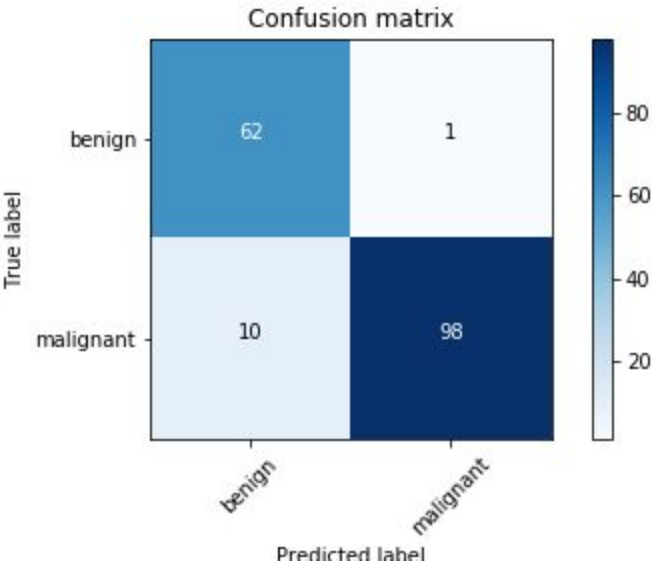
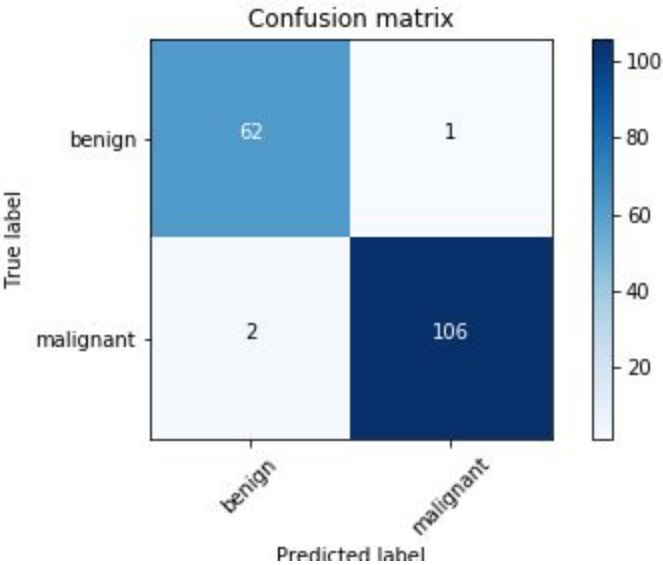


Fig. 1: On breast cancer (SLP)



**Fig. 2: On breast cancer (MLP
with 1 hidden layer)**

Fig. 3: On breast cancer (SLP)

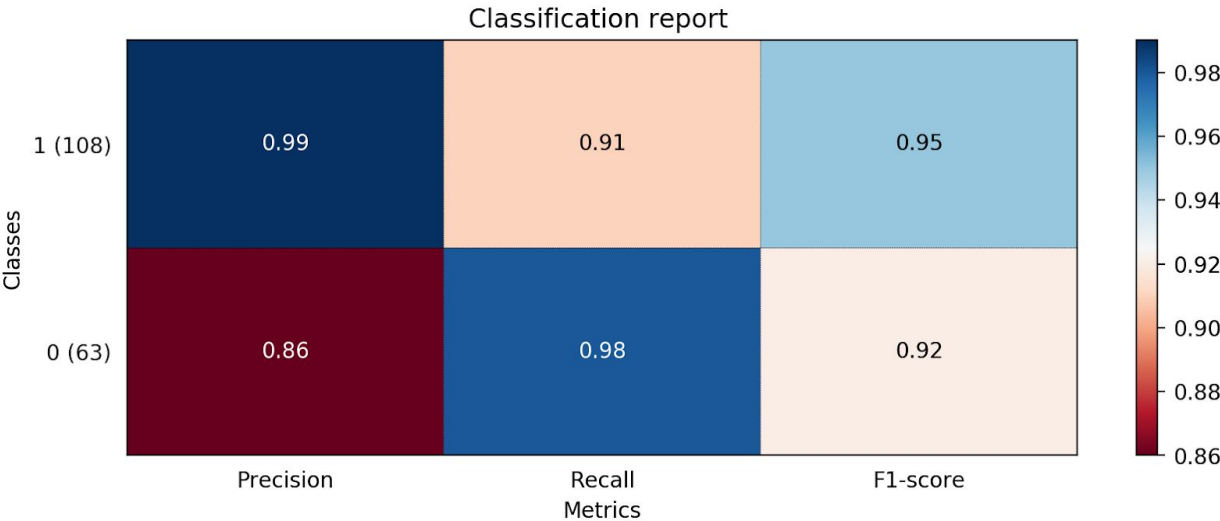


Fig. 4: On breast cancer (MLP with 1 hidden layer)

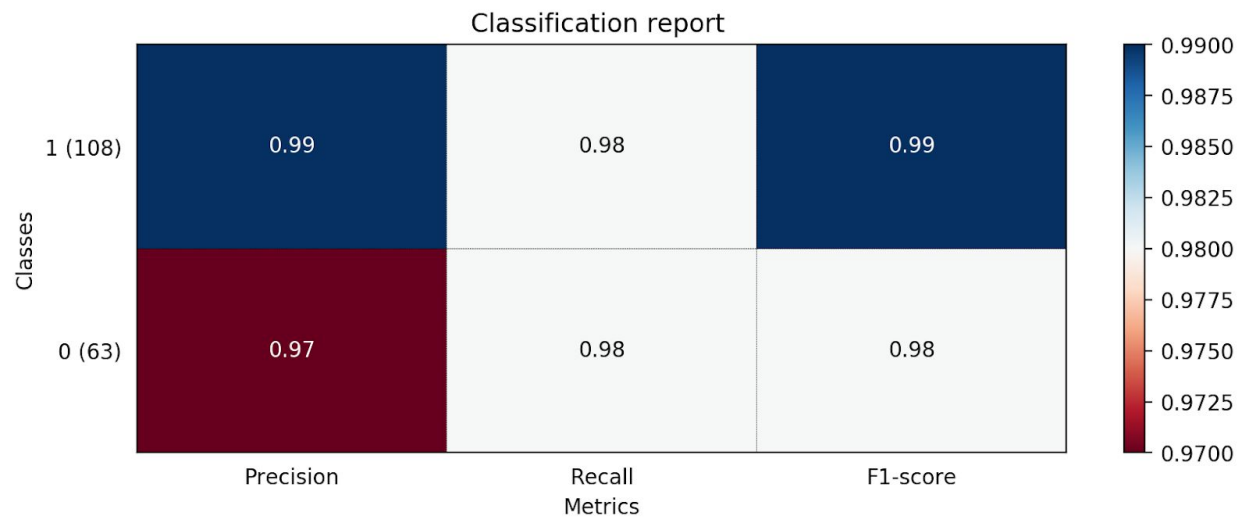


Fig. 5 : Accuracy comparisons of SLP and MLP with 1 hidden layer on breast-cancer dataset

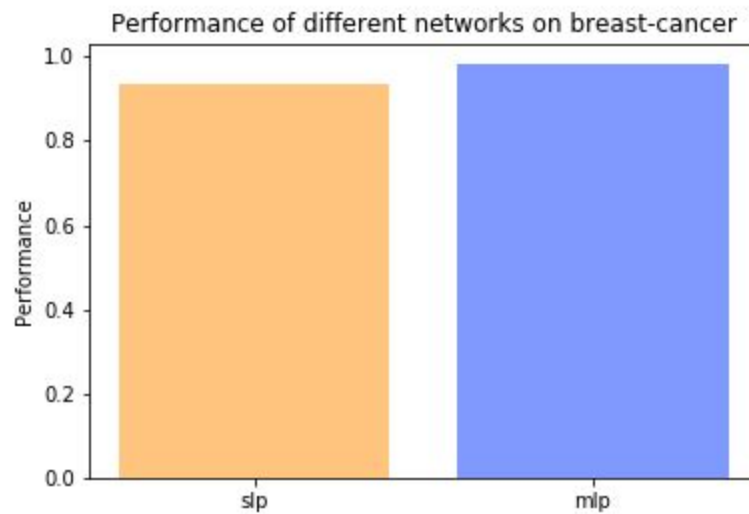
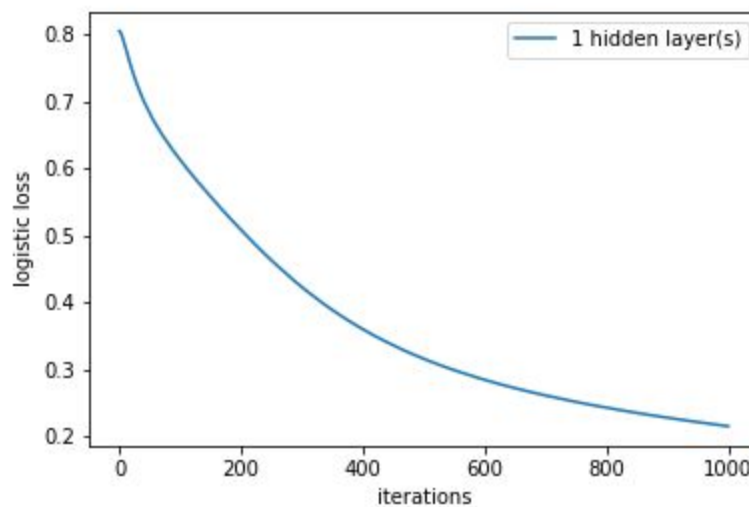


Fig. 6: Logistic loss plot on breast-cancer dataset



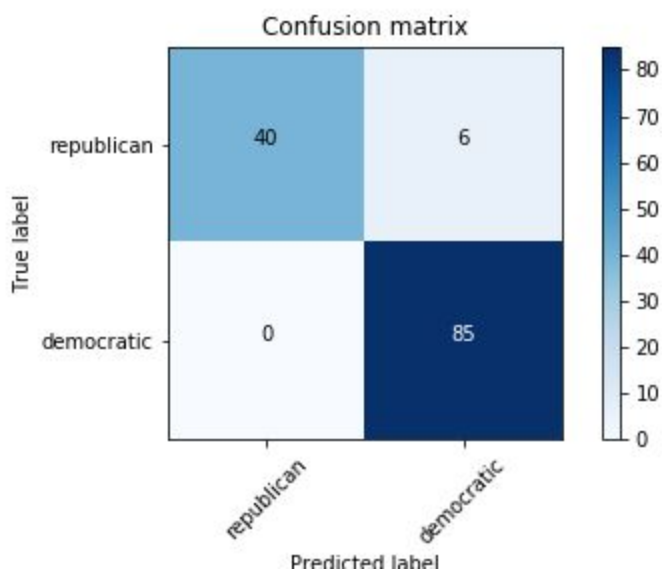


Fig. 7: On house votes 84 (SLP)

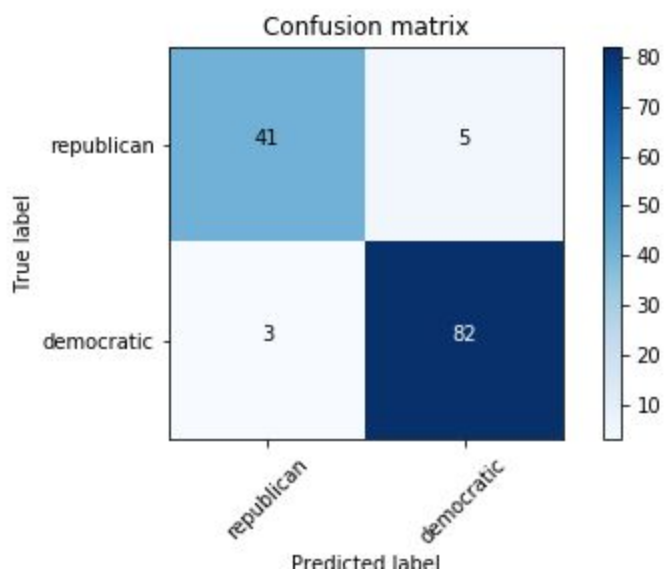


Fig. 8: On house votes 84 (MLP with 1 hidden layer and SGD optimizer)

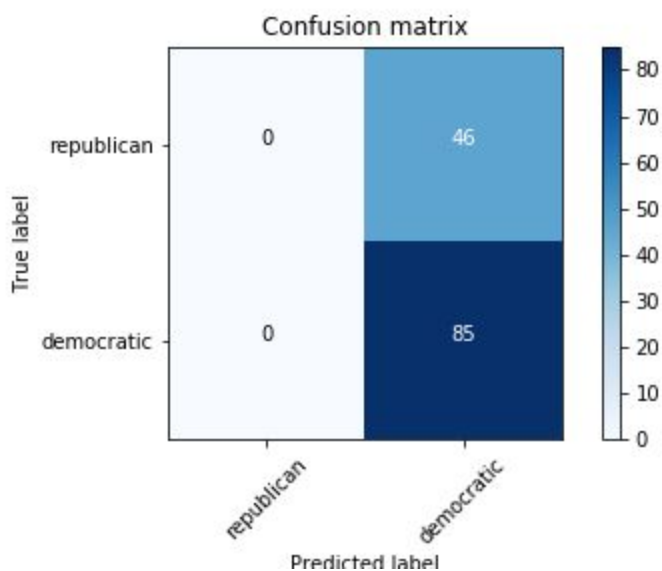


Fig. 9: On house votes 84 (MLP with 2 hidden layer and SGD optimizer)

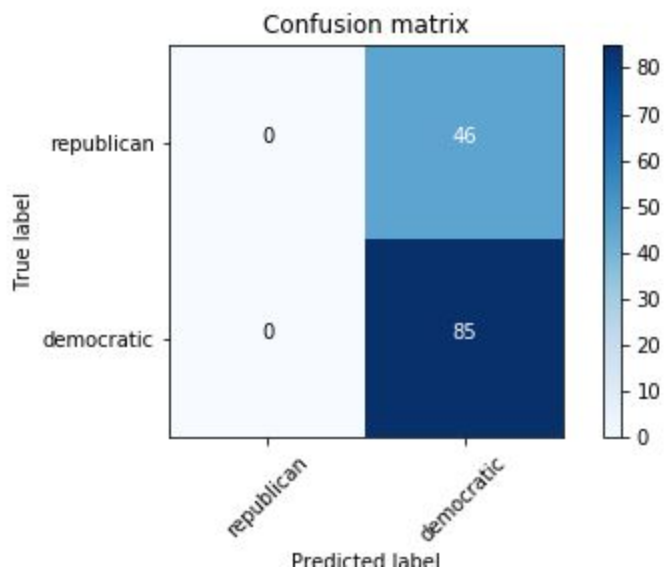


Fig. 10: On house votes 84 (MLP with 3 hidden layers and SGD optimizer)

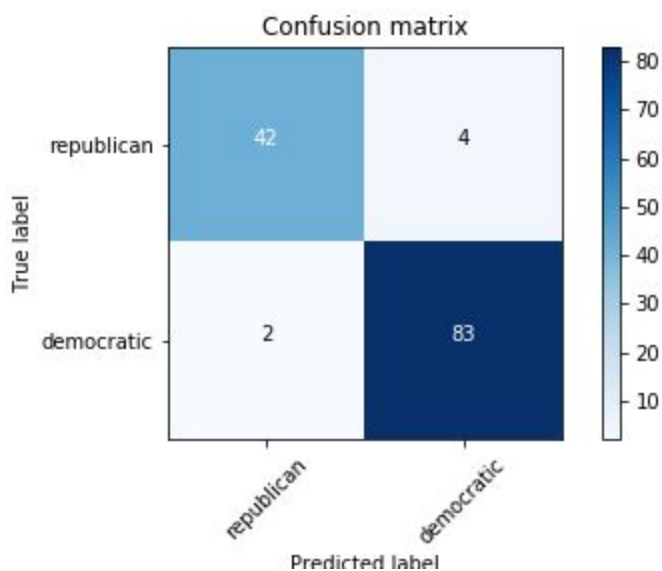


Fig. 11: On house votes 84 (MLP with 1 hidden layer and LBFGS)

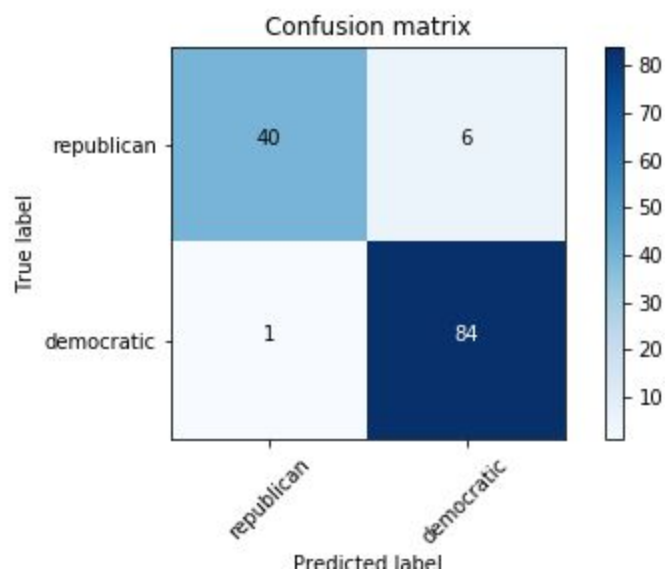


Fig 12: On house votes 84 (MLP with 2 hidden layers and LBFGS)

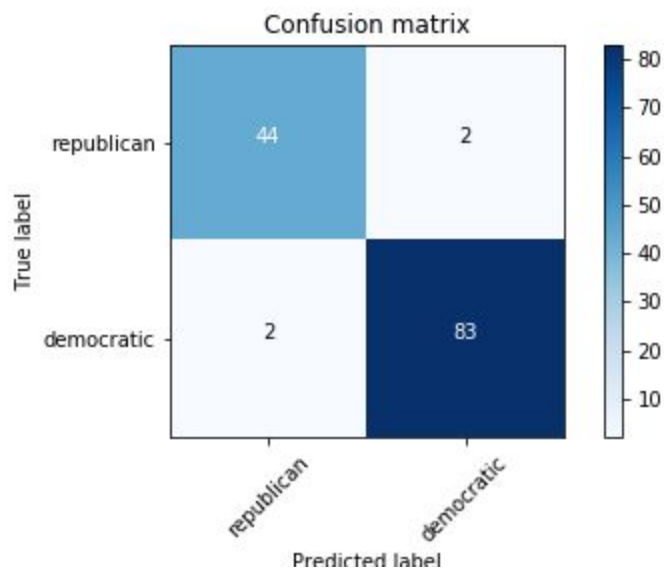


Fig. 13: On house votes 84 (MLP with 3 hidden layers and LBFGS)

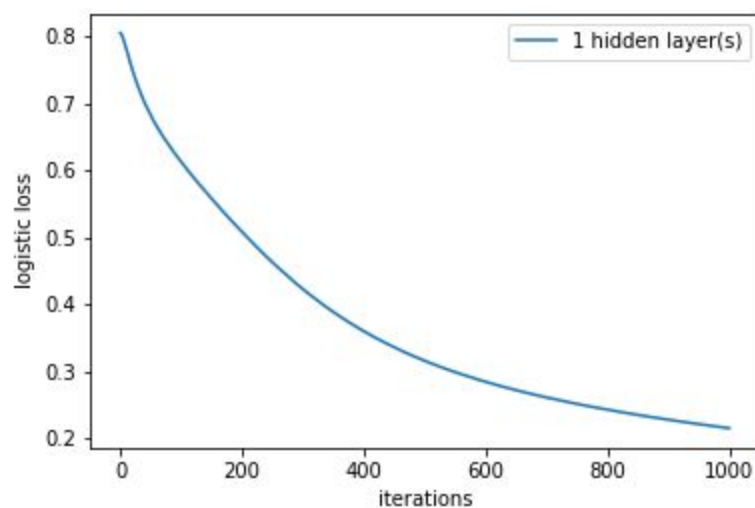


Fig. 14: On house votes 84 (logistic loss plot of MLP with 1 hidden layer and SGD)

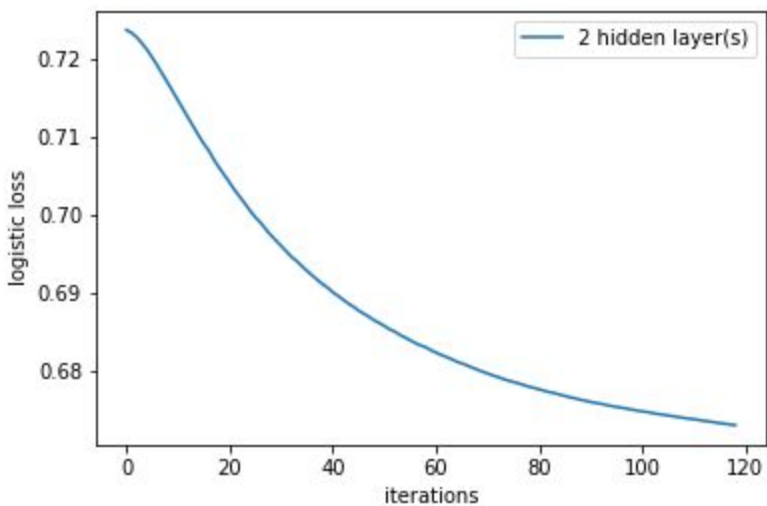


Fig 15: On house votes 84 (logistic loss plot of MLP with 2 hidden layers and SGD)

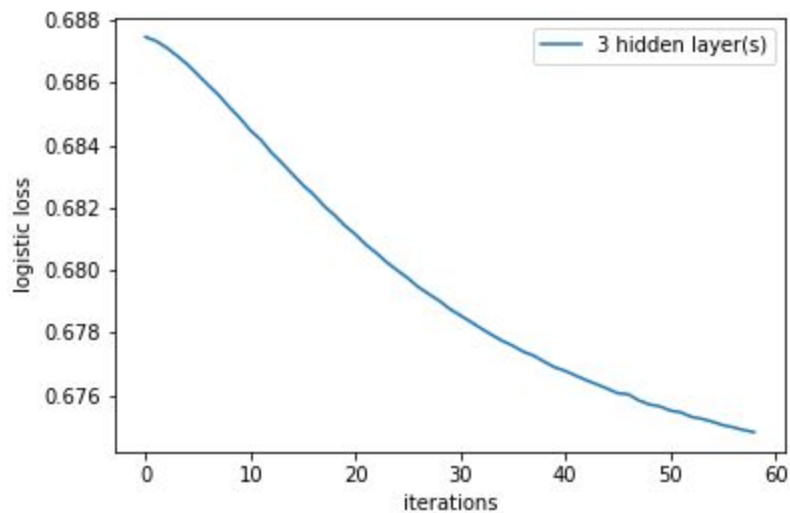


Fig 16: On house votes 84 (logistic loss plot of MLP with 3 hidden layers and SGD)

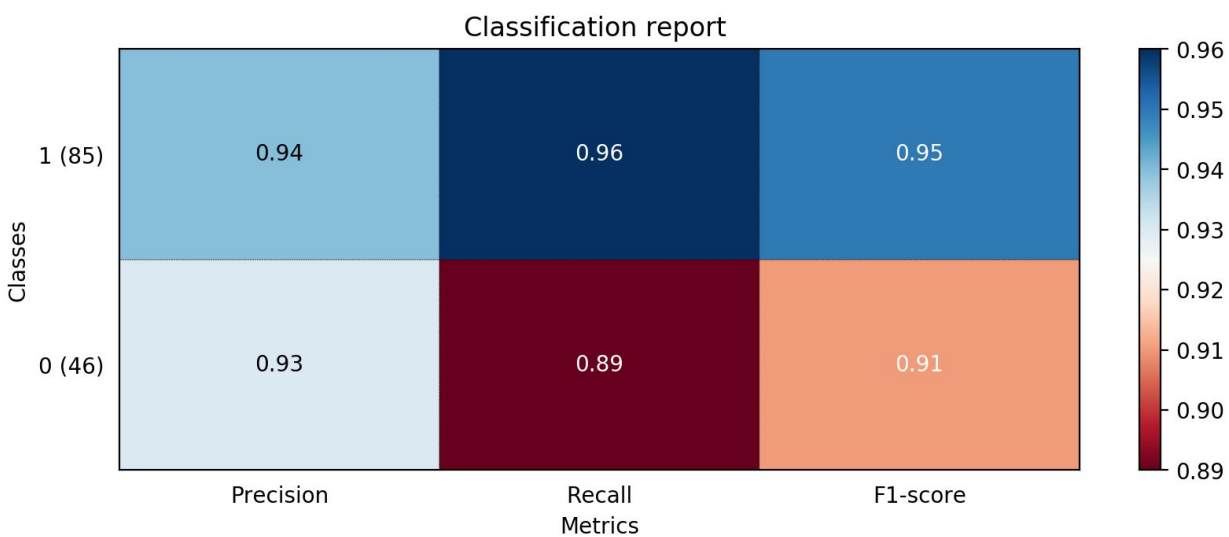


Fig. 17: On house votes 84 (MLP with 1 hidden layer and SGD)

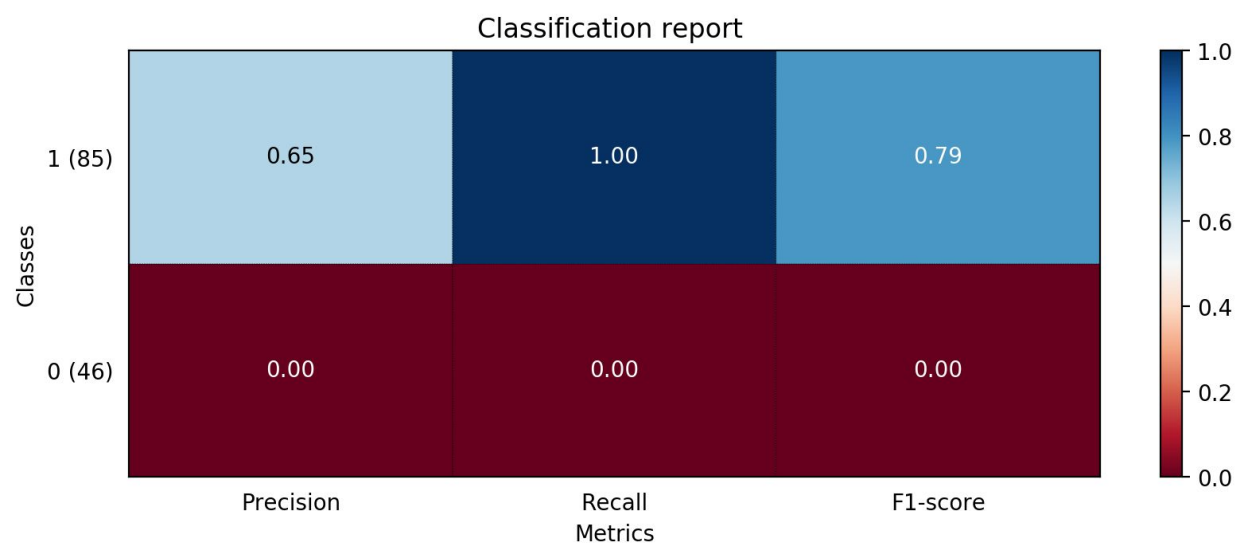


Fig. 18: On house votes 84 (MLP with 2 hidden layers and SGD)

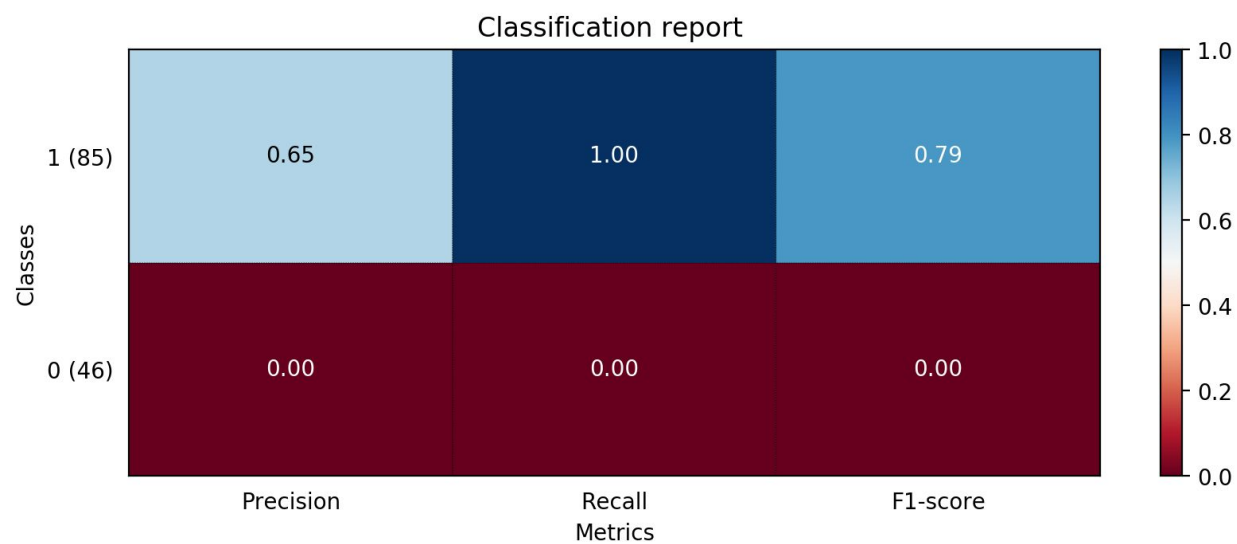


Fig. 19: On house votes 84 (MLP with 3 hidden layers and SGD)

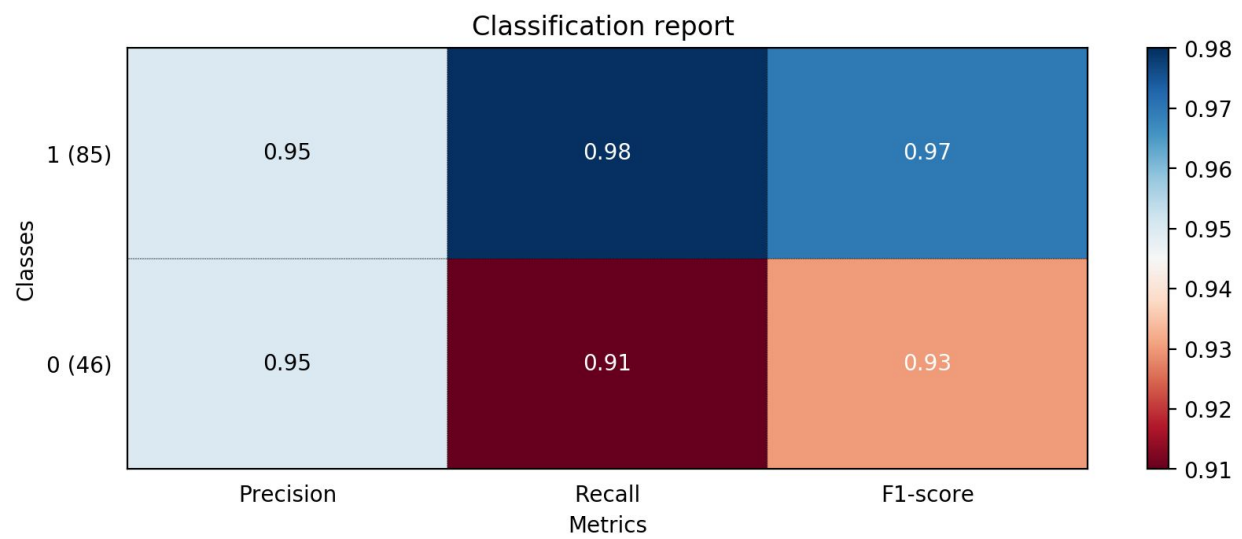


Fig. 20: On house votes 84 (MLP with 1 hidden layer and LBFGS)

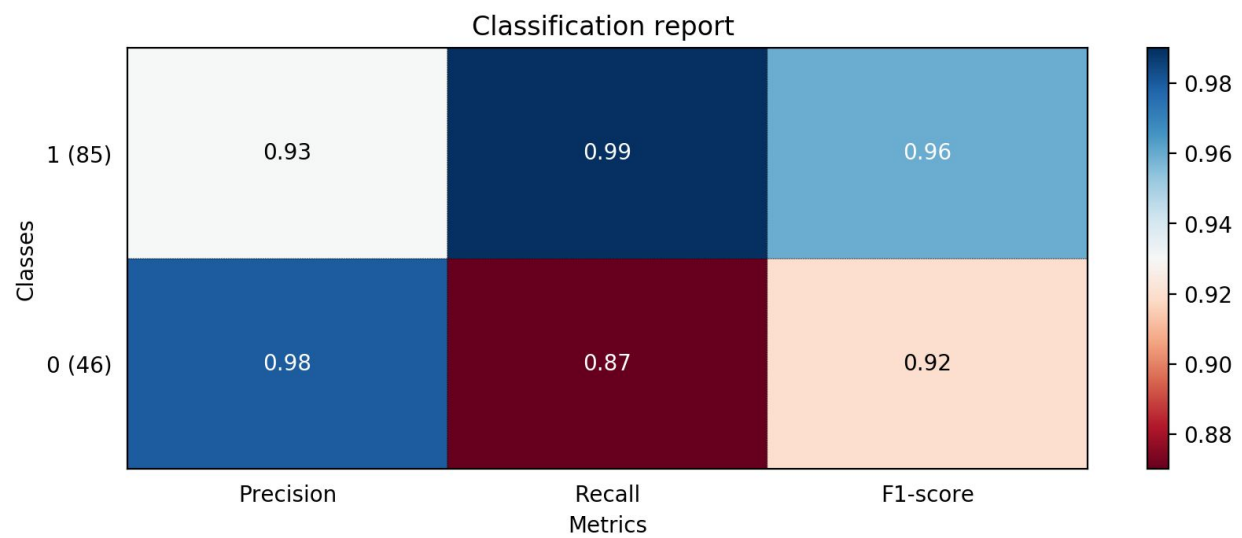


Fig. 21: On house votes 84 (MLP with 2 hidden layer and LBFGS)

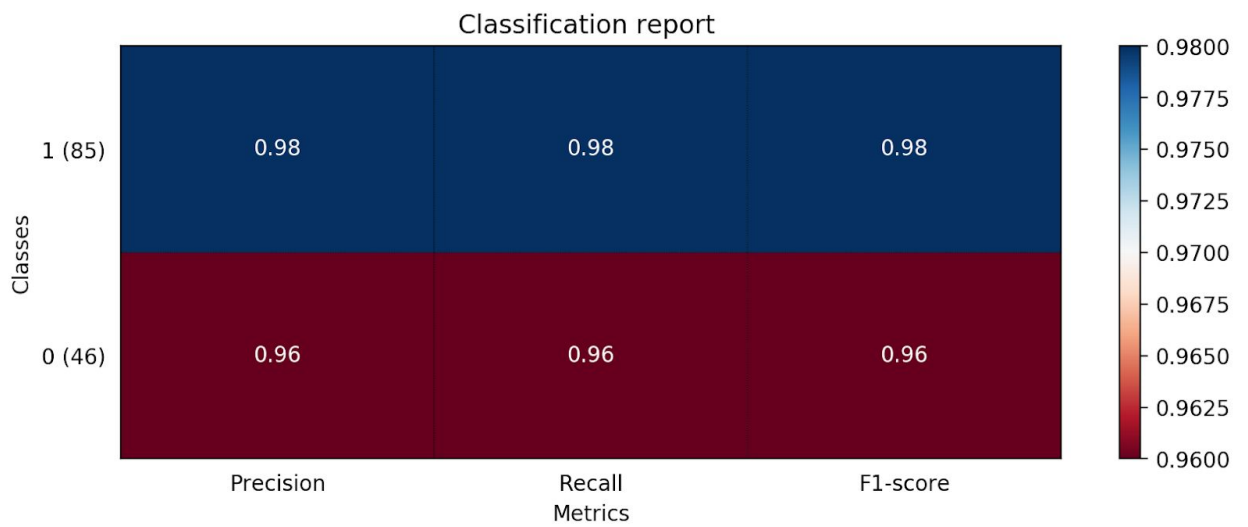


Fig. 22: On house votes 84 (MLP with 3 hidden layers and LBFGS)

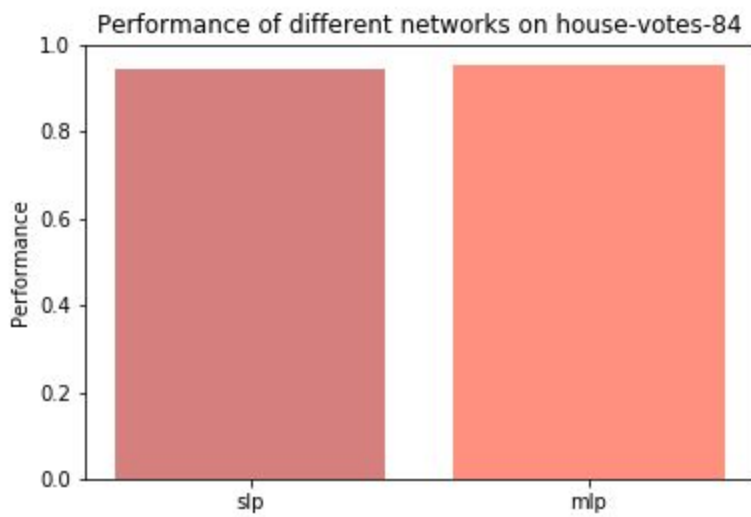


Fig. 23: On house votes 84 (SLP vs MLP with LBFGS)

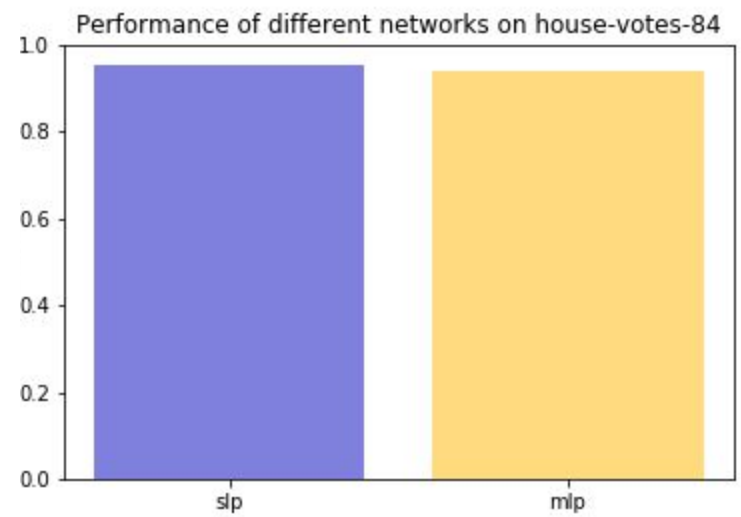


Fig. 24: On house votes 84 (SLP vs MLP with SGD)

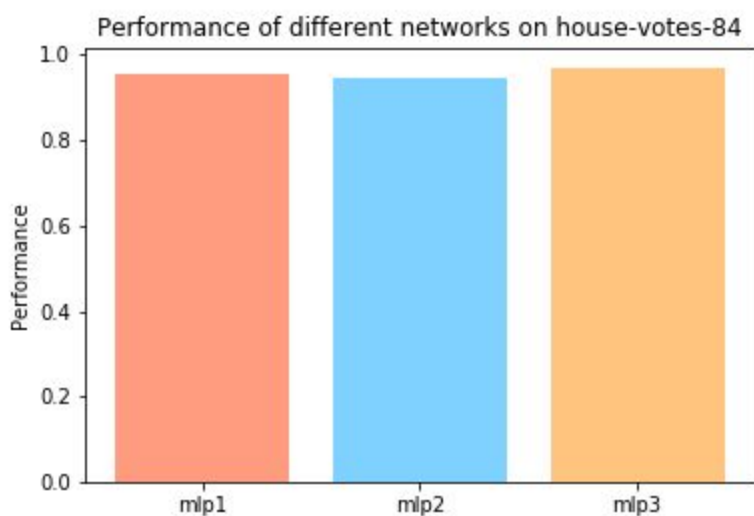


Fig. 25: On house votes 84 (MLP with 1, 2 and 3 hidden layers and LBFGS)

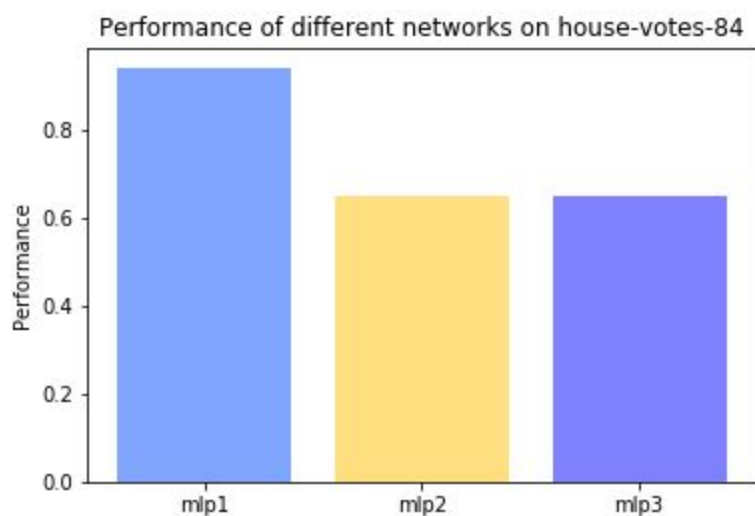


Fig. 26: On house votes 84 (MLP with 1, 2 and 3 hidden layers and SGD)

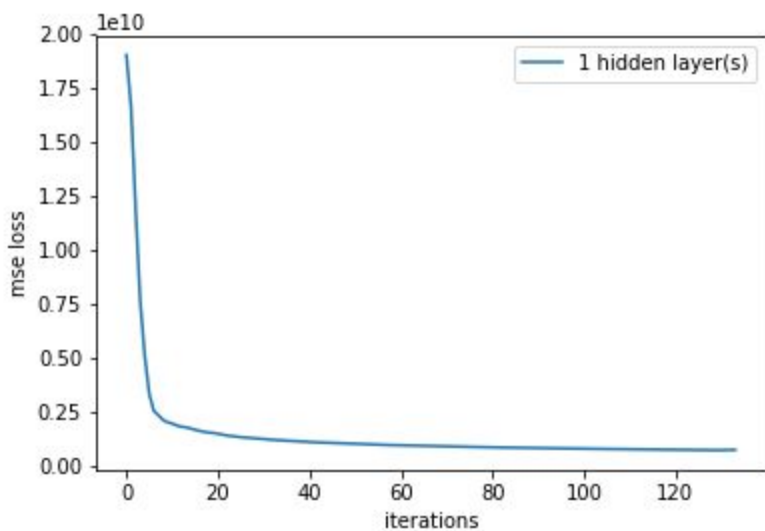


Fig. 27: MSE loss plot on house price (MLP with 1 hidden layer)

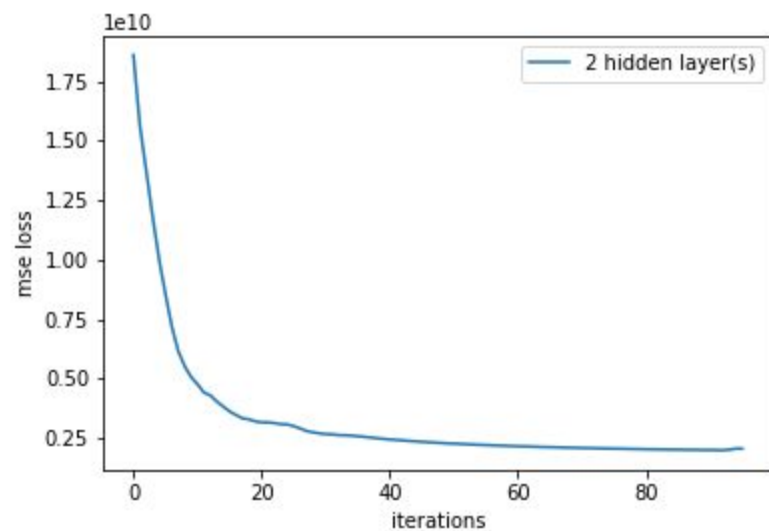


Fig. 28: MSE loss plot on house price (MLP with 2 hidden layers)



Fig. 29: Means of error (on house price)

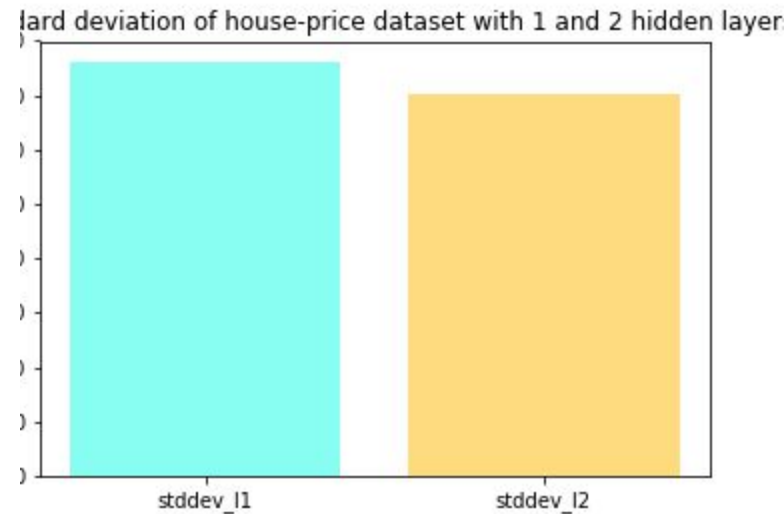


Fig. 30: Standard deviations of error (on house prices)

❖ CONCLUSIONS

- SLPs and MLPs with 1 hidden layer have almost the same performances for the given datasets.
- For small datasets, LBFGS converges faster than SGD as an optimizer.
- As the number of layers in an MLP increases, SGD takes lesser number of iterations to minimize the loss.

NAME: SANDIPAN SARMA
ENROLLMENT ID: 510515076