Name(s): Girish Madhavan Venkataramani
NetID(s): gmv4
Team name on Kaggle leaderboard: Erased_GM

Note : Since the hyper parameter arrays were small enough and my training procedure was sequential updates to the hyperparameters, I had run through the combinations manually by hand instead of a iterative structure. So you might not find the outputs for other combinations of hyper parameters in the python notebook

Perceptron

For both datasets, changing the learning rate values among [0.4, 0.5, 0.6, 0.7, 0.8, 0.9], 0.5 gave the best training accuracy. Rice dataset being small and low dimensional, low number of epochs like 10 was more than enough for the training accuracy to converge, But for the Larger Fashion-MNIST dataset more epochs like 50 were needed for the accuracy to settle. Epoch values tried : [10, 20, 30, 40, 50, 60], both cases. Addition of learning rate decay allowed the weights to converge and kept it from changing, allowing better accuracies. Low decay rate like 0.95 didn't make much of a difference from not using it, but using something like aggressive like 0.6 or 0.7 was good enough to freeze the model at good accuracy

RICE DATASET

Optimal hyperparameters:	Learning rate = 0.5 Learning rate decay = 0.6 Number of Epochs = 10 Batch size = 64
Training accuracy:	98.121 %
Validation accuracy:	98.047 %
Test accuracy:	98.185 %

Fashion-MNIST DATASET

Optimal hyperparameters:	Learning rate = 0.5 Learning rate decay = 0.6 Number of Epochs = 50 Batch size = 512
Training accuracy:	84.430 %
Validation accuracy:	82.810 %
Test accuracy:	82.200 %

SVM

RICE Dataset, it needed a higher value of regularization constant and more number of epochs to converge. For tuning, first the learning rate decay was varied between [0.7, 0.76, 0.78, 0.8, 0.82] to get the best accuracy, followed by changing the number of epochs [10, 15, 20, 25, 30, 35, 40] to get the best result. However for the Fashion MNIST dataset, a lower regularization constant and higher number of epochs gave better accuracy. Without those changes the Fashion MNIST dataset was stuck at 78% accuracy

RICE DATASET

Optimal hyperparameters:	Learning rate = 0.6 Learning rate decay = 0.8 Batch size = 64 Number of Epochs = 25 Regularization Constant = 0.5
Training accuracy:	98.149 %
Validation accuracy:	97.773 %
Test accuracy:	98.158 %

Fashion-MNIST DATASET

Optimal hyperparameters:	Learning rate = 0.76 Learning rate decay = 0.8 Batch size = 1024 Number of Epochs = 35 Regularization Constant = 0.05
Training accuracy:	84.520 %
Validation accuracy:	82.650 %
Test accuracy:	82.110 %

Softmax

For the RICE dataset : Just setting a low regularization constant value increased the training accuracy, the values tried were [0.01, 0.04, 0.1, 0.25, 0.5, 0.9, 1]. The number of epochs were increased till the training accuracy stops changing, values tried were [18, 20, 25, 30, 35, 45], out of which 45 epochs seemed excessive.
For the Fashion-MNIST dataset : Continuing from the previous case, the batch size was first fixed from [512, 1024, 2048] with the option giving the highest training accuracy. Followed by increasing the temperature and slightly increasing the Epsilon value to 0.01, values tried [0.01, 0.02, 0.1, 0.2] to find the model that allowed me to cross 80% accuracy on the training dataset. The final threshold accuracy was crossed by using iterating through the following values for learning rate : [0.1, 0.2, 0.3, 0.4, 0.5 …1] This helped in finding the model that gave an accuracy of close to 86 % in the training dataset

RICE DATASET

Optimal hyperparameters:	Learning rate = 0.9 Learning rate decay = 0.8 Batch size = 64 Temperature = 1 Epsilon [label smoothing] = 0 Number of Epochs = 35 Regularization constant = 0.01 Seed [weight initialization] = 0
Training accuracy:	98.020 %
Validation accuracy:	97.828 %
Test accuracy:	98.075 %

Fashion-MNIST DATASET

Optimal hyperparameters:	Learning rate = 0.1 Learning rate decay = 0.76 Batch size = 2048 Temperature = 25 Epsilon [label smoothing] = 0.01 Number of Epochs = 25 Regularization constant = 0.5 Seed [weight initialization] = 18
Training accuracy:	85.900 %
Validation accuracy:	84.770 %
Test accuracy:	83.540 %

Logistic

Slowly changing learning rate decay by increments of 0.2 starting from 0, till the training accuracy reaches its maximum in a fixed number of epochs. Similarly the batch size used in mini batch can be determined, going with batch size = 1 led to very low accuracy due to the drastic changes in the gradient. High batch size like 512 or 1024, gave better accuracy than 1 but it was still not as good as using 64 as the batch size. From this we should consider using smaller batch sizes for smaller data sets [approximately 10% of the size of training dataset]. Once the other hyperparameters have been tuned, we can iterate through seed values 0 – 200 or more to find the weight initialization that gives the best accuracy. All this is possible because the dataset being small and its relatively low dimensionality as compared to the Fashion-MNIST dataset.

RICE DATASET

Optimal hyperparameters:	Learning rate = 0.5 Learning rate decay = 0.6 Batch size = 64 Number of Epochs = 10 Threshold = 0.5 seed [weight initialization] = 67
Training accuracy:	98.221 %
Validation accuracy:	98.103 %
Test accuracy:	98.323 %