CSc 84020 Neural Networks and Deep Learning
Homework 2: Classification and Prediction with DL Multi-Layer Perceptron
Andrea Ceres
Shao Liu

3b. Analysis

To set up our experiments, we split our dataset into approximately 70% train, 20% validation, and 10% test sets, with each set remaining balanced among the six classes. We did this by splitting off 10% for the test set, and 22% (20/90) of the remaining for validation. This analysis covers use of the optimizer Stochastic Gradient Descent with Nesterov Momentum (SGD) and the activation function Rectified Linear Unit (ReLU).

Twelve combinations were run covering permutations of dropout rates {0.0, 0.1, 0.2, 0.4} and L2 regularization {0.0, 0.0005, 0.01}. The resulting validation accuracy and loss are shown in *Table.1*.

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Dropout	0.0	0.1	0.2	0.4	0.0	0.1	0.2	0.4	0.0	0.1	0.2	0.4
L2 Regularization	0.0	0.0	0.0	0.0	0.0005	0.0005	0.0005	0.0005	0.01	0.01	0.01	0.01
Accuracy	**0.7644**	0.7634	0.7527	0.7245	**0.7618**	0.7553	0.7495	0.7060	0.7211	0.7111	**0.7249**	0.6897
Loss	**0.6877**	0.6928	0.7090	0.8019	**0.7508**	0.7567	0.7769	0.9156	0.9464	0.9581	**0.9387**	1.0390

Table.1

From the descriptive statistics and analysis report, we determined a benchmark test accuracy of 0.6717 using K-Nearest Neighbors (KNN) on a small subset of our dataset. All experiments of Multilayer Perceptrons surpassed this accuracy, with the highest test accuracy of 0.7644. Counterintuitively, our base MLP model (without dropout and without L2 regularization) yielded the best results—the highest accuracy and the lowest loss.

The higher the dropout introduced, the poorer the results. This is particularly evident for 40% dropout. The base model with its 29,046 trainable parameters appears to underfit the data. Therefore, the introduction of dropout is inappropriate, as it is normally used to counteract overfitting.

Furthermore, the higher the L2 regularization introduced, the poorer the results. Similar to the use case for dropout, L2 regularization helps generalize the model to new data by reducing overfitting. This is predicated on an overfit model. In the case of our base model, the underfitting model topology does not warrant regularization prior to further architectural adjustments.

The validation accuracy and validation loss followed the training accuracy and training loss most closely in the cases of dropouts 0.1 and 0.2. For each of our experiments, validation accuracy and loss remained fairly flat after the first five epochs. The loss curve failing to decrease is an indication that our models underfit our dataset.

Reflected in the precision-recall curves and consistent in almost all normalized confusion matrices and classification reports, the *mosquito* class resulted in the lowest recall, while the *butterfly* class gave the highest recall. Given the poorer performance with accuracy and loss, the models with dropout and L2 regularization also yielded slightly lower AUC scores and worse ROC curves.