

CSCE 636: Homework 1

Aryan Sharma
UIN: 326006767

Question 1

1a. The function `train_valid_split`, splits the dataset into training and validation sets. The training dataset is the actual dataset that we use to train the model. However, if we use the entire training dataset, we can overfit and may not generalize well to the test data. The validation dataset is thus used to tune the hyperparameters. By testing on a set of validation examples that the models were not trained on, we obtain a better estimate of each hypothesis h_i 's true gene

1b. We can use the full dataset to produce our final model as the more data we use the more likely it is to generalise well but we should also make sure that we obtain an unbiased performance estimate via nested cross-validation and potentially consider penalising the cross-validation statistic to further avoid over-fitting in model selection.

1c. Implemented in code

1d. This is a bias addition in the Weight matrix.

1e. Implemented in code

1f. Shown in Fig. 1

Question 2

1a. Implemented in code

1b. Implemented in code

1c. Shown in Fig. 2

1d. Implemented in code. The accuracy is —.