Assignment 1 Solution

1. Training set: A training dataset is a dataset of examples used during the learning process and is used to fit the parameters of the neural network.

   Validation set: A validation dataset is a dataset of examples used to tune the hyperparameters of the neural network.

   Test dataset: A test set is a set of examples used only to assess the performance of the neural network.

2. (1). Define the dataset by using tensor.
   (2). Define the neural network model.
   (3). Configure the learning process.
   (4). Training the network with the fit() function.

3. Number of units: Positive integer, dimensionality of the output space.

   Activation: Activation function to use.

4. Optimizer: a learning algorithm called gradient descent, refers to the calculation of an error gradient or slope of error and “descent” refers to the moving down along that slope towards some minimum level of error.

   Loss: compute the quantity that a model should seek to minimize during training.

   Metrics: list of metrics to be evaluated by the model during training and testing.

   Difference between loss and metrics: The loss function is used to optimize your model. The metric is used to judge the performance of your model.

5. Input: the data feed in to the neural network.

   Output: the target data that the neural network wants to match.

   Batch Size: a hyperparameter of gradient descent that controls the number of training samples to work through before the model’s internal parameters are updated.

   Epoch: a hyperparameter of gradient descent that controls the number of complete passes through the training dataset.

   Validation data: Data on which to evaluate the loss and any model metrics at the end of each epoch.

6. (1) convert text string to word index list.
   (2) encode list as binary vectors indicating whether each word is present or not.

7. Underfitting: occurs when a statistical model cannot adequately capture the underlying structure of the data.

   Overfitting: the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably.

8. Layers: the number of layers, type of layers.

   Units per layer: dimensionality of the output space.

   Activation functions: different type of activation functions.

   Loss: different type of loss functions.

9. If $1/(1+e^{x})$ >= 0.5, output = 1; else output = 0.

10. The class label is mapped to integer values. Then, each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1.

11. The output of the softmax function can be used to represent a categorical distribution – that is, a probability distribution over K different possible outcomes.

12. Use the Categorical Crossentropy loss function when there are two or more label classes. We expect labels to be provided in a one hot representation. If you want to provide labels as integers, we use Sparse Categorical Crossentropy loss. There should be number of classes floating point values per feature.

13. 20%.

14. x = (x - mean)/std. Prevent features with wider ranges from dominating the distance metric.

15. $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$ $\qquad$ $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
    MAE is easier to interpret.

16. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k − 1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used

exactly once as the validation data. The k results can then be averaged to produce a single estimation.

The k-fold cross-validation is needed when evaluate model on a limited data sample.

17. Taking the average of the k validation errors. To train the final model, we use the entire training data.