|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number | Training | Validation | Test | Performance on Test Set (Loss, Accuracy) |
| Model 2 | 100 | 10000 | 5000 | (0.692, 0.593) |
| Model 3 | 1000 | 10000 | 5000 | (0.529, 0.737) |
| Model 4 | 25000 | 10000 | 5000 | (0.317, 0.898) |
| **Model 5** | **35000** | **10000** | **5000** | **(0.268, 0.925)** |

**Report**

**Objective:**

The goal is to develop an embedding model suitable for training on the IMDB reviews dataset. This model should discern the features within the reviews and then classify whether a given review is positive or negative.

**Varied Sample Sizes: (Scratch Models)**

**Outcome:**

Initially, Models 2 and 3 underwent training on a limited dataset utilizing a basic network structure comprising solely an embedding layer. While the model's performance wasn't notably poor, it also didn't excel.

In contrast, Models 4 and 5 incorporated Convolution 1D alongside embedding and dense layers to evaluate potential performance enhancements. This adjustment yielded several insights.

Firstly, the expansion of the training dataset enabled the model to glean more insights about the data, particularly regarding the positive and negative aspects of reviews, thereby facilitating improved generalization on the test set. Secondly, integrating Conv1D with the embedding layer bolstered the model's resilience by allowing it to assimilate more information about reviews as filters traversed through them, leading to more accurate predictions on the test set.

During the model construction phase, fine-tuning hyperparameters such as embedding vector dimensions, learning rate, Conv1D layers, dense layers, dropout rate, and node configurations proved pivotal in enhancing the model's ability to generalize. Plotting the training and validation loss/accuracy aids in identifying the optimal point where the model begins to overfit. Moreover, adopting a strategy of constructing a complex network architecture to intentionally induce overfitting and subsequently fine-tuning parameters has proven most effective in achieving superior generalization on the test set. Lastly, in instances of poor model performance, augmenting the dataset size has consistently proven beneficial for improving learning capacity.

**Final evaluation:**

Based on the provided information, it's evident that the fifth model emerged as the strongest performer among the five models created. This particular model underwent training using 15,000 samples for training, 1,000 samples for validation, and 5,000 samples for testing.

A graph with red dots

Description automatically generated with low confidence

A picture containing text, screenshot, display, number

Description automatically generated

**Pre-trained network:**

The vector dimension of the Glove network was used to build these models and the network constitutes LSTM layers.

**Outcomes:**

In the realm of constructing machine learning models, underfitting poses a more pressing concern than overfitting. When a model underfits, it signifies poor performance on the training data, indicating even worse performance on unseen data. Models 3 and 4 serve as examples of underfit models, exhibiting subpar performance on the training set and even poorer performance on the test set.

Occasionally, employing a straightforward architecture with fewer input layers and nodes can yield superior performance compared to complex architectures. Model 5, characterized by a simple network architecture, avoided underfitting and demonstrated commendable performance on the training set. Consequently, its performance on the test set surpassed that of the other models.

**Final Evaluation:** Judging from the evaluation outcomes, it's evident that the fifth model stood out as the top performer among the five constructed models. This particular model underwent training with 15,000 samples, validation with 1,000 samples, and testing with 5,000 samples.

A graph with red dots

Description automatically generated with low confidenceA graph with blue dots

Description automatically generated with low confidence