Report: Neural Network model Architecture and model evaluation for SMS Classification

Model Architecture

The neural network model implemented for SMS classification consists of the following layers and configurations:

Input Layer:

- Input dimension: 5000 (Number of TF-IDF features)
- Normalized using StandardScaler.

Hidden Layers:

1. **First Hidden Layer:**
    - Layer Type: Dense
    - Number of Neurons: 128
    - Activation Function: ReLU (Rectified Linear Unit)
    - Additional Components: Dropout with a rate of 0.3 to prevent overfitting.
2. **Second Hidden Layer:**
    - Layer Type: Dense
    - Number of Neurons: 64
    - Activation Function: ReLU
    - Additional Components: Dropout with a rate of 0.3 to prevent overfitting.

Output Layer:

- Layer Type: Dense
- Number of Neurons: 1
- Activation Function: Sigmoid (for binary classification).

Compilation

- **Optimizer:** Adam (Adaptive Moment Estimation) for efficient gradient-based optimization.
- **Loss Function:** Binary Crossentropy, suitable for binary classification tasks.
- **Metrics:** Accuracy.

Model Training

- **Epochs:** 10
- **Batch Size:** 32
- **Validation Split:** 20% of the training data reserved for validation.

Summary of Model Layers

Layer (Type)	Output Shape	Parameters
Dense (Input)	(None, 128)	640,128
Dropout	(None, 128)	0
Dense	(None, 64)	8,256
Dropout	(None, 64)	0
Dense (Output)	(None, 1)	65

Model Evaluation

The model achieved the following performance metrics on the test data:

Class/Metric	Precision	Recall	F1-Score
0	0.98	0.99	0.99
1	0.92	0.89	0.91
Accuracy	-	-	0.97
Macro Avg	0.95	0.94	0.95
Weighted Avg	0.97	0.97	0.97

Experimental Observations

- Dropout significantly reduced overfitting during training.
- ReLU activation in hidden layers helped in faster convergence and handling non-linear patterns.
- Sigmoid activation in the output layer ensured predictions between 0 and 1 for binary classification.

Conclusion

The neural network architecture designed for SMS classification demonstrated robust performance with an overall accuracy of 97%. Key architectural choices, such as the use of Dropout layers to mitigate overfitting and the ReLU activation function for efficient learning, proved effective. The model's precision and recall scores indicate its reliability in identifying both spam and non-spam messages.

Future work could explore:

- Optimizing hyperparameters for improved performance.
- Expanding the dataset for better generalization.
- Testing alternative architectures, such as LSTM or GRU, for sequential data handling.

This architecture highlights the practicality of neural networks in solving binary classification problems with high dimensional input features.