Nicholas Ilvento

DSE 6211 Final Report

Executive Summary

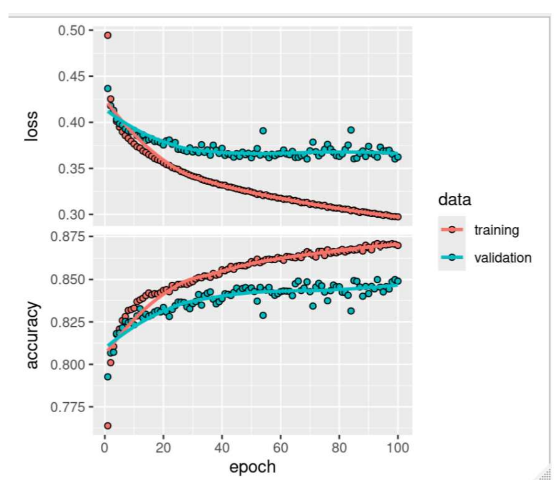
Project Goal: ABC Hotels would like to analyze the probability of their guests canceling their reservation. To help address this business need, a variety of machine learning techniques will be applied as means to build a sufficient neural network. The goal of this network is to create an accurate model in which we can identify reservations and/or guests that have a high risk of cancellation. These high-risk clusters will then be the focus of ABC Hotel's marketing efforts in order to obtain as much business as possible, create more opportunities for revenue, and prevent future canceled reservations.

Key Findings:

We developed two neural network models to predict reservation cancellations. The first model has an architecture with two hidden layers consisting of 50 and 25 units. While this was our initial model proposed, there was much trial and error that went into determining this architecture. Nonetheless, we created a second model with a deeper architecture. This was for the purpose of creating overfitting and more accurate results. The second model has three hidden layers comprising 800, 600, and 200 units. While this model was slightly more accurate, we are not 100% convinced it's the most ideal selection for ABC Hotels. Depending on the processing power of ABC Hotels, the initial model proved may be a more efficient option compared to the second model. While there is a miniscule drop off in accuracy shown by our AUC Values, the initial model is much faster to run and can create cancellation predictions faster. Depending on the speed at which ABC Hotels would like to analyze their bookings. I believe we have created two great networks and should give the company an option of which one they'd like.

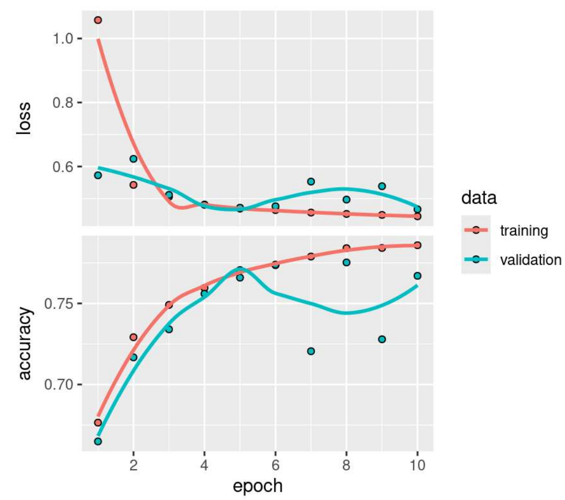
Initial Model:

```
model <- keras_model_sequential(list(
  layer_dense(units =50, activation = "relu"),
  layer_dense(units = 25, activation = "relu")
  ,layer_dense(units = 1, activation = "sigmoid")
))

compile(model,
        optimizer = "rmsprop",
        loss = "binary_crossentropy", metrics = "accuracy")


history <- fit(model, training_features, training_labels,
               epochs = 100, batch_size = 100, validation_split = 0.33)
plot(history)
```

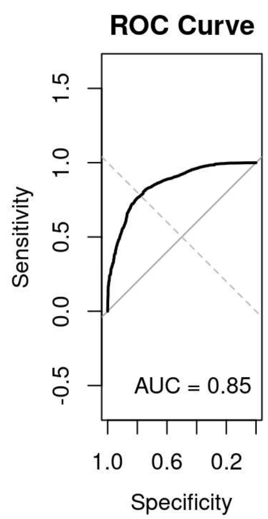


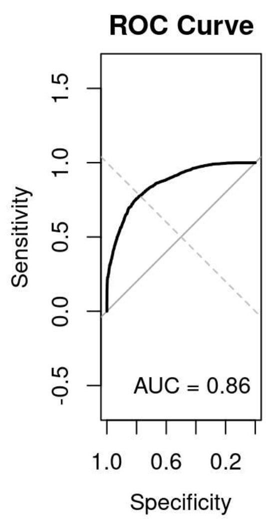
Second Model:

```
model2 <- keras_model_sequential(list(
  layer_dense(units =800, activation = "relu"),
  layer_dense(units =600, activation = "relu"),
  layer_dense(units = 200, activation = "relu"),
  layer_dense(units = 1, activation = "sigmoid")
))

compile(model2,
        optimizer = "rmsprop",
        loss = "binary_crossentropy", metrics = "accuracy")


history2 <- fit(model2, training_features, training_labels,
               epochs = 10, batch_size = 100, validation_split = 0.33)
plot(history2)
```



Despite increasing the number of hidden layers and units within these layers, a significantly more accurate overfit was unable to be achieved. Many different architectures were experimented with, and all were less accurate or around as accurate as our initial model.





As shown above the ROC Curves and AUC values are extremely similar between both models. Considering our initial model was experimented with, we do trust the analysis it provides. The initial model is also much faster than our second model since it is less complex. This would provide ABC Hotels with a more efficient way of predicting cancellations.

Nicholas Ilvento

Recommendations:

We recommend integrating the initial or final model into the reservation system for real-time predictions. This will allow ABC Hotels to use predictions to prioritize high-risk bookings for targeted marketing and special offers. Along with this, future analyses should consider incorporating additional data features and further model tuning to enhance the model's accuracy and performance.

Approach & Data

Overall Approach: The goal of this project was to predict reservation cancellations at ABC Hotels. This would enable the hotel to identify high-risk bookings and take proactive measures to reduce cancellations, ultimately optimizing revenue and occupancy rates.

Available Data: The dataset provided included various features related to bookings, excluding the 'Booking_ID' which was not significant for the neural network.

Feature Engineering Steps:

1. Converted 'booking_status' to a binomial format (1 = canceled, 0 = not_canceled).
2. Transformed 'arrival_date' to extract 'month' and 'season'.
3. One-hot encoded categorical features like 'season', 'type_of_meal_plan', 'room_type_reserved', and 'market_segment_type'.
4. Scaled numerical features.

Machine Learning Methods Used:

- Activation functions: ReLU for hidden layers, sigmoid for the output layer.
- Loss function: Binary Cross-Entropy Loss.
- Optimization algorithm: RMSprop.

Detailed Findings and Evaluation

Initial Model:

- Architecture: Two hidden layers (50 and 25 units).
- Activation functions: ReLU & sigmoid
- Performance: Achieved 85% accuracy.
- Evaluation: ROC curve and AUC analysis showed good performance, but potential underfitting.

- Much faster to run than our additional model.

Additional Model:

- Architecture: Three hidden layers (100, 50, and 25 units).
- Performance: Improved flexibility and higher potential for overfitting.
- Evaluation: ROC curve and AUC analysis showed slightly improved performance.
- Performance: Achieved 86% accuracy.

ROC Curves and AUC Scores:

- The ROC curves for both models indicated good discriminatory power, with the additional model showing a slightly higher AUC score, reflecting a more accurate, but slower performance.

Recommendations

Model Usage in Practice:

- Integrate the selected model into ABC Hotels' reservation system for real-time assessment of booking cancellation risks.
- Utilize the predictions to prioritize high-risk bookings for targeted marketing efforts and special offers.
  - An example of this could be rewarding reservations with low risk of cancellations. This can help build loyalty and satisfaction among hotel guests.
- Depending on the time-frame of ABC Hotels, both models can be utilized. Our initial model may be better for our clients when ABC Hotels needs to create accurate predictions on the spot. If ABC Hotels has time to spare and high processing power, the second model can be used to obtain slightly more accurate results.

Future Research/Steps Forward:

- Consider incorporating additional data features such as more customer demographics and booking history for more comprehensive analysis.
- Further experiment with different neural network architectures and hyperparameters to continue improving model performance and create the desired overfit with over 90% accuracy.
- Explore other machine learning techniques like ensemble methods for potential improvements.