# Tripadvisor Las Vegas Hotel Reviews - Neural Network Sentiment Analysis

Kayson Hansen - kayson@stanford.edu, Emily Chiao - emchiao@stanford.edu

## Predicting

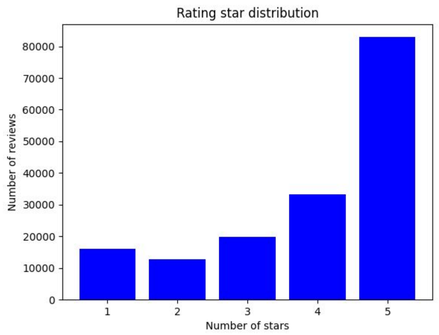
We built binary and multi-class classification neural network models to conduct sentiment analysis on Tripadvisor hotel reviews in Las Vegas.
Our motivation was to better understand the type of language people use when describing their opinions on experiences and investigate whether reviews for casinos are generally positive or negative.

The inputs to our algorithm were paragraph embedding vectors generated from the hotel reviews using doc2vec. The outputs are either a positive (1) or negative (0) sentiment for the binary classification model or a prediction of how many stars (1-5) for the multi-class classification model.

We were able to achieve **90.28%** test accuracy for the neural network binary output and **60.47%** test accuracy for the neural network multi-class output.

## Data

We scraped approximately 150,000 reviews from Tripadvisor for our data. This included the text of the review and a corresponding star rating (1-5) attached to the review. The graph below shows the rating distribution of the reviews we scraped.




## Features

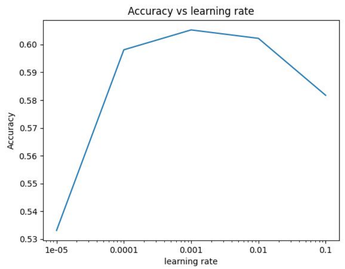
The features for our data were the entries in the 50-dimensional vectors generated by doc2vec. doc2vec takes in documents as inputs (in our case, the Tripadvisor hotel reviews) and outputs vectors that attempt to capture the theme or overall meaning of the documents. These features were appropriate for our task because they allowed us to encapsulate the sentiment of our reviews in a numerical format that we could then use to train our models.

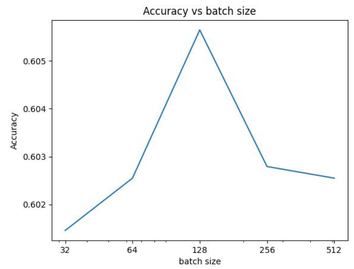
## Models

**Neural network (multi-class output)**
This model had 3 hidden layers (48, 20, 10 units) and a softmax output layer.
**Baseline: logistic regression (multi-class output)**
This model had a softmax output layer.
Softmax $\sigma(z)_i = e^{z\,i} / \sum_{j=1}^{K} e^{z\,j}$

**Neural network (binary output)**
This model had 3 hidden layers (48, 20, 10 units) and a sigmoid output layer.
**Baseline: logistic regression (binary output)**
This model had a sigmoid output layer.
Sigmoid $f_{w,b}(x) = 1/(1 + e^{-z})$ where $z = w \times x^{(i)} + b$

The figures below show the results of training our multi-class neural network model with various learning rates and batch sizes. We found that the optimal learning rate was **0.001**, and the optimal batch size was **128**.




## Results

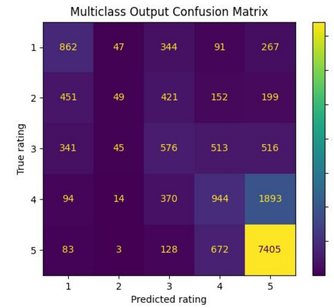
We conducted stratified sampling with our data to obtain a train set with 80% of the data, a cross-validation set with 10%, and a test set with 10%. The table below represents the error for each of our datasets on the models that we built. We found the cross-validation error to be similar to the train error, indicating that our models had low variance.

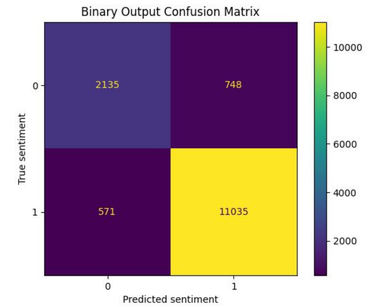
| Model | Train Error | CV Error | Test Error |
|---|---|---|---|
| Neural network (multi-class) | 39.21% | 39.77% | 39.53% |
| Logistic regression (multi-class) | 40.80% | 41.13% | 40.55% |
| Neural network (binary) | 8.87% | 9.21% | 9.72% |
| Logistic regression (binary) | 11.07% | 11.32% | 10.96% |

## Discussion

Overall, our results were closely in line with what we expected. We achieved an error rate of **only 9.72%** when we classified the sentiment of reviews as positive or negative, and we achieved a larger, but still, reasonable, error rate of **39.53%** when we classified reviews as a certain number of stars. In both cases, our neural network models outperformed the corresponding logistic regression models—which we used as baseline models to compare our performance to—with error rates around **1%** lower.

The confusion matrices below show our model's performance on each type of review. Our model with binary outputs had an error rate of **4.92%** on positive sentiment reviews and an error rate of **25.95%** on negative sentiment reviews. We attribute this disparity to there being many more positive reviews than negative reviews on Tripadvisor. For the model with multiclass outputs, the most obvious problems are misclassifying **57.10%** of the 4-star reviews as 5-star and misclassifying **35.46%** of the 2-star reviews as 1-star and **33.10%** as 3-star, both of which seem like mistakes humans would make as well.




## Future

If we had another 6 months to work on this, we would first get more data. We would scrape reviews from hotels in other locations besides Las Vegas and also look into scraping reviews in other categories on Tripadvisor besides hotels (e.g. restaurants). This would allow us to get a broader view of online reviews and how well the sentiment of reviews can be predicted based on text.

Secondly, we would try training different types of machine learning models: decision trees, k-means clustering, etc. In our limited time for this project, we were only able to train neural network and logistic regression models, so it would be beneficial to investigate the effectiveness of other types of models.

## References

"The 10 Best Hotels in Las Vegas, NV for 2023." *Tripadvisor*, https://www.tripadvisor.com/Hotels-g45963-Las_Vegas_Nevada-Hotels.html.

"Doc2vec paragraph embeddings." *Gensim*, https://radimrehurek.com/gensim/models/doc2vec.html