CS6140 Final Project: Airline Passenger Satisfaction Investigation

Group 2: Christopher Daly, Jacqueline Girouard, Edgar Alan David

**Introduction**

As the world slowly emerges from the Covid-19 pandemic, travel is reentering the lives of millions of Americans. How can airline companies capitalize on this return to travel and make the experience as positive as possible so industry can quickly recover? We are going to use machine learning, particularly the support vector machines (SVM) method, to explore what makes an airline trip positive. What matters most to travelers and what relationships do these factors have to one another? These are some of the general questions that we would like to answer to know what satisfaction means to airline passengers.

**The Idea**

We will be implementing a support vector machine on airline passenger satisfactions. We will implement various kernels to explore various possible decision boundaries and will explore how feature selection affects results.

**Dataset**

We will be utilizing the public “Airline Passenger Satisfaction” dataset from Kaggle. It features 25 possible features and with over 1500 data entries. We chose this dataset because it is likely some features are only negligibly important, and it will be interesting to explore how we are able to reveal this. We will be striving to predict the “satisfaction” feature, which is binary: either “satisfied” or “neutral or dissatisfied”.

Link: <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

**Project Plan**

**Tools and Methodology:**

We will use the scikit-learn functions to implement an SVM with linear, polynomial and radial-basis-function kernels. The scikit-learn SVM implementation allows for custom kernels. As part of our experimentation, we can explore additional decision boundaries if time permits. We will use tools such as pandas and NumPy to handle data normalization and manipulation.

**Visualizations:**

We will use matplotlib and seaborn as our visualization tool. We will be visualizing our decision boundary, noting the false positives, true positives, false negatives and true negatives. We will use the seaborn correlation-matrix to do this. We will also strive to provide curves demonstrating the accuracy of our models as feature selection becomes sparser (how do we do with only 2 vs 3 features)?

**Regularization Techniques:**

We will adjust the C (regularization parameter) in the SVM to allow for varying hyperplanes to account for potentially noisy data. We will also be mindful of the gamma parameter which allows for reducing the number of points influencing the boundary.

**Project Timeline**

Below are detailed tasks to be completed each week.

Week 9 (July 4th):

*Establish shared code for experiments.*

1. Create a shared git repository for collaborative development (Edgar)
2. Normalize Data Features where necessary (Christopher)
3. Implement training, test split (Christopher)
4. Implement a testing function to assess the accuracy, precision, recall and f1 score of future results. (Jacqueline)
5. Visualization of Correlation Matrix (Edgar)

Week 10 (July 11th):

*Experiment on different kernels on various features*

1. Implement an SVM with a linear kernel. Explore how feature reduction improves results (Edgar)
2. Implement an SVM with a polynomial kernel. Explore how feature reduction improves results (Jacqueline)
3. Implement an SVM with a radial-basis-function kernel. Explore how feature reduction improves results (Christopher)

Week 11 (July 18th): Results Analysis.

*Are we able to identify features that are not useful or potentially noisy? Can our results be confirmed by a different implementation? For example, if our experiments reveal a polynomial decision boundary, a linear decision boundary model should not be able to perform as well.*

1. Visualize results (Christopher)
2. Implement alternate linear model. (Edgar)
3. Compare results with alternate linear model. (Jacqueline)

Week 12 (Julu 25th): Formalize paper.

1. Write Final Report (Team)
2. Create and Practice Final Presentation (Team)
3. Generate Final Video (Team)

**References**

Chunyu Chen, Xiaohui Qi and Maoliu Lin, "Research and application of noise suppression based on support vector machine," IEEE International Symposium on Communications and Information Technology, 2005. ISCIT 2005., 2005, pp. 358-361, doi:

10.1109/ISCIT.2005.1566868.

Xu H., Caramanis C., Mannor S., (2009). Robustness and Regularization of Support Vector Machines. Journal of Machine Leaning Research 10.

https://www.jmlr.org/papers/volume10/xu09b/xu09b.pdf