JAWS Team Project

William Langley, Alex Cummins, Jakob Ramirez, Saron Abebe

Data Description

- Each school has five *data points* attributed to them.
  - State, County, Title 1 eligibility, Total students, and Total reduced-price lunch eligible students
- Data organized by year, ranging from 2008 to 2017.
- *Raw data includes null values*

normalize total_students and number of students on reduced_price lunch

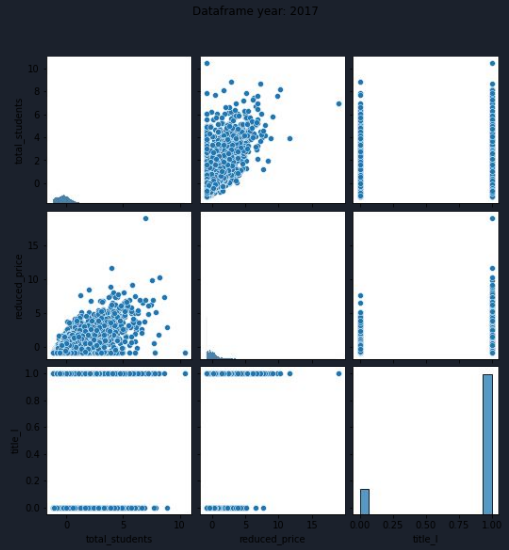
```
realcols = ["total_students","reduced_price"]
for k,_df in dates_df.items():
    for col in realcols:
      mean = _df[col].mean()
      std = _df[col].std()
      _df[col] = (_df[col]-mean)/std
```

Questions Moving Forward

- How can we predict a school's title status based on the data provided?
- How should we predict this result? Classification or regression?
- Which data points should we analyze to find our answer?

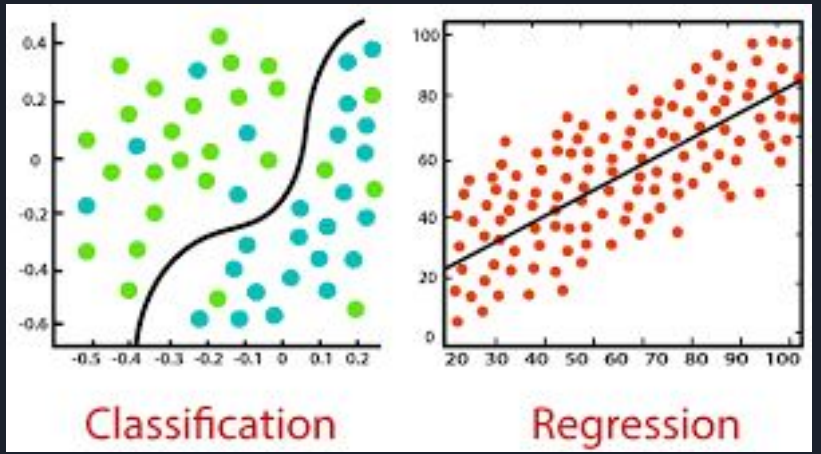
Feature Selection

- Reduced Price Lunch is the best estimator of Title 1 status



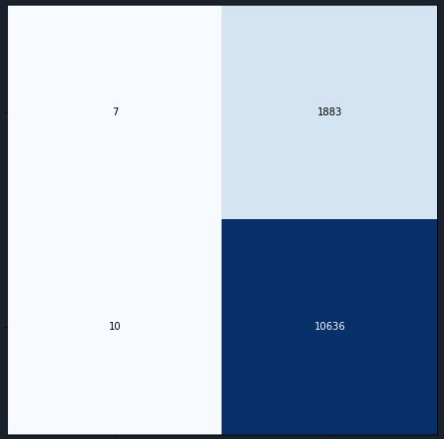
Classification vs Regression

- Classification for predicting a label and regression is used for predicting quantity
- We are trying to determine whether schools will have the Title 1 label, therefore we should use classification

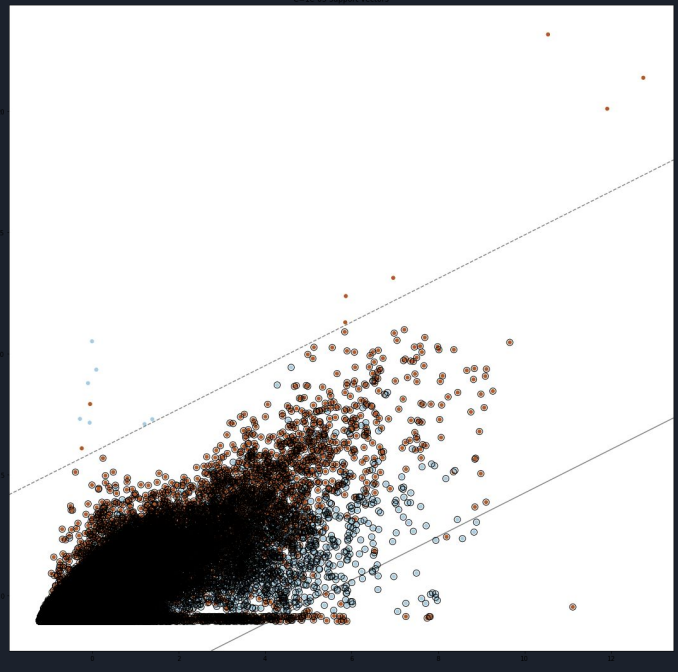


Support Vector Machine

- Used to separate the two classes of data points, the goal is to find the plane with the maximum margin
- Score: 0.8489948947032546



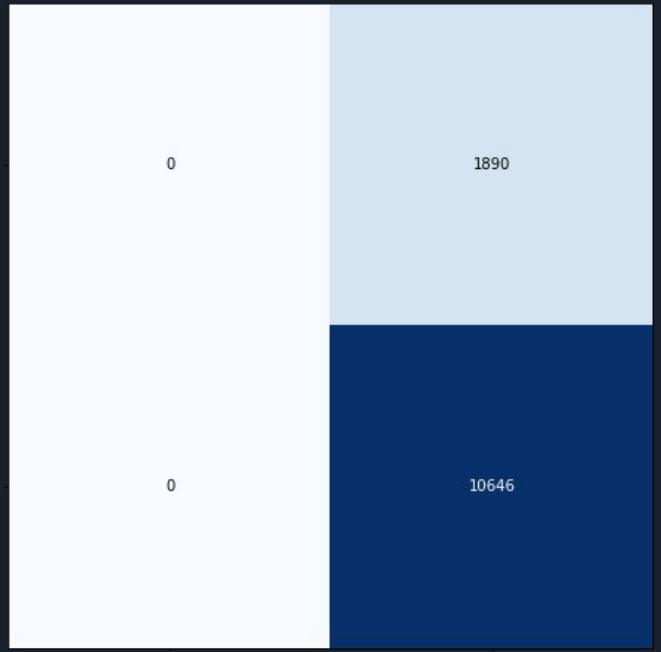
SVM Confusion Matrix



SVM Plot

Stochastic Gradient Descent

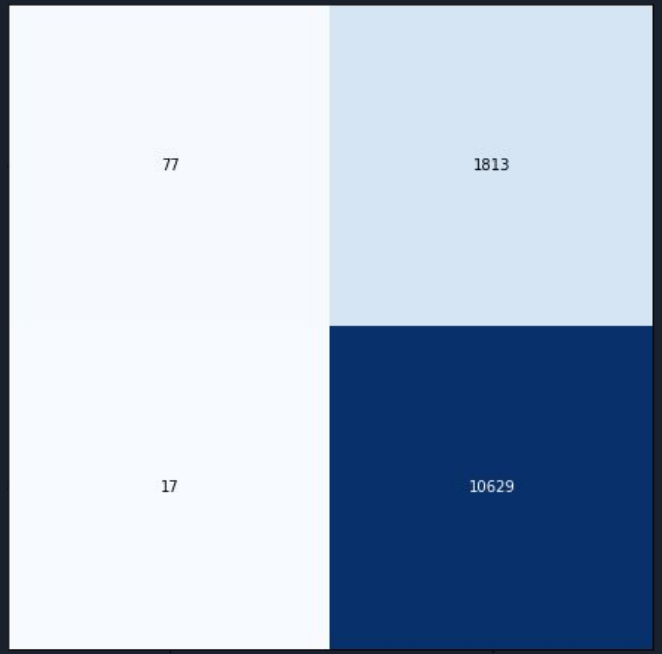
- Stochastic Gradient Descent is a iterative method of approximation for a gradient descent optimization.
- Gradient descent seeks to find minimums by moving in the opposite direction of the gradient
- Score: 0.849234205488194



SGD Confusion Matrix

Decision Tree

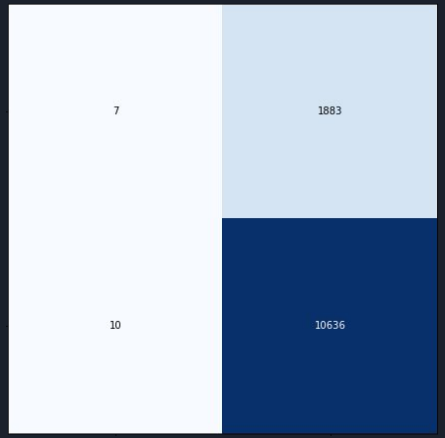
- A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label
- Score: 0.8540204211869815



Decision Tree Confusion Matrix

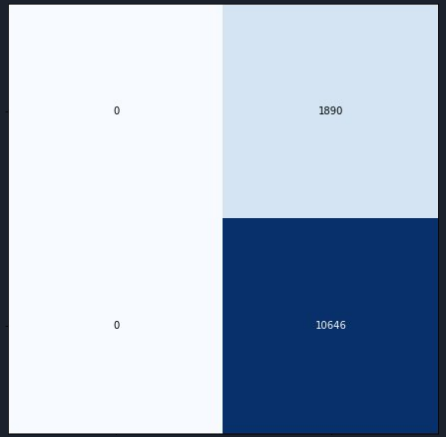
Summary of Results

Support Vector Machine



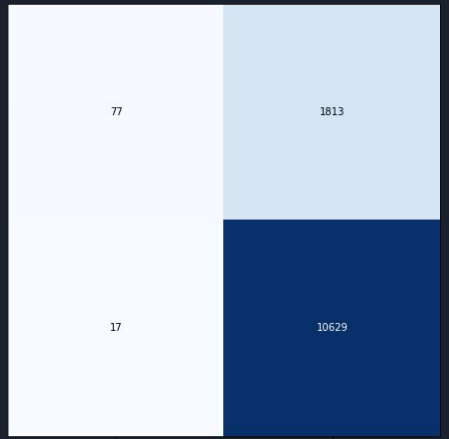
Score: 0.848994894703

Stochastic Gradient Descent



Score: 0.8492342054881

Decision Tree



Score: 0.8540204211869

Best Model

Conclusion

- Total Students in the school and reduced price lunch are good indicators of title 1 status.
- Our three models (Support Vector Machine, Stochastic Gradient Descent, and Decision Tree) all had similar results but the decision tree was slightly better.
- Questions / Feedback?