

Project 3

Predicting Credit Card Default

Cheng Ji

Content

- The Data
- EDA
- Mine
- Refine
- Model
- Conclusion

The Data

- Credit Card default info.
- Goal: predict the probability of default based on personal and previous payment information.

EDA and Mining

- Clean dataset, no missing values, appropriate data types.
- Rename columns for easy reference.

EDA and Mining

- Plot bill amount variables and payment amount variables. Strong Correlation among those variables, which indicates feature selection.

Refining

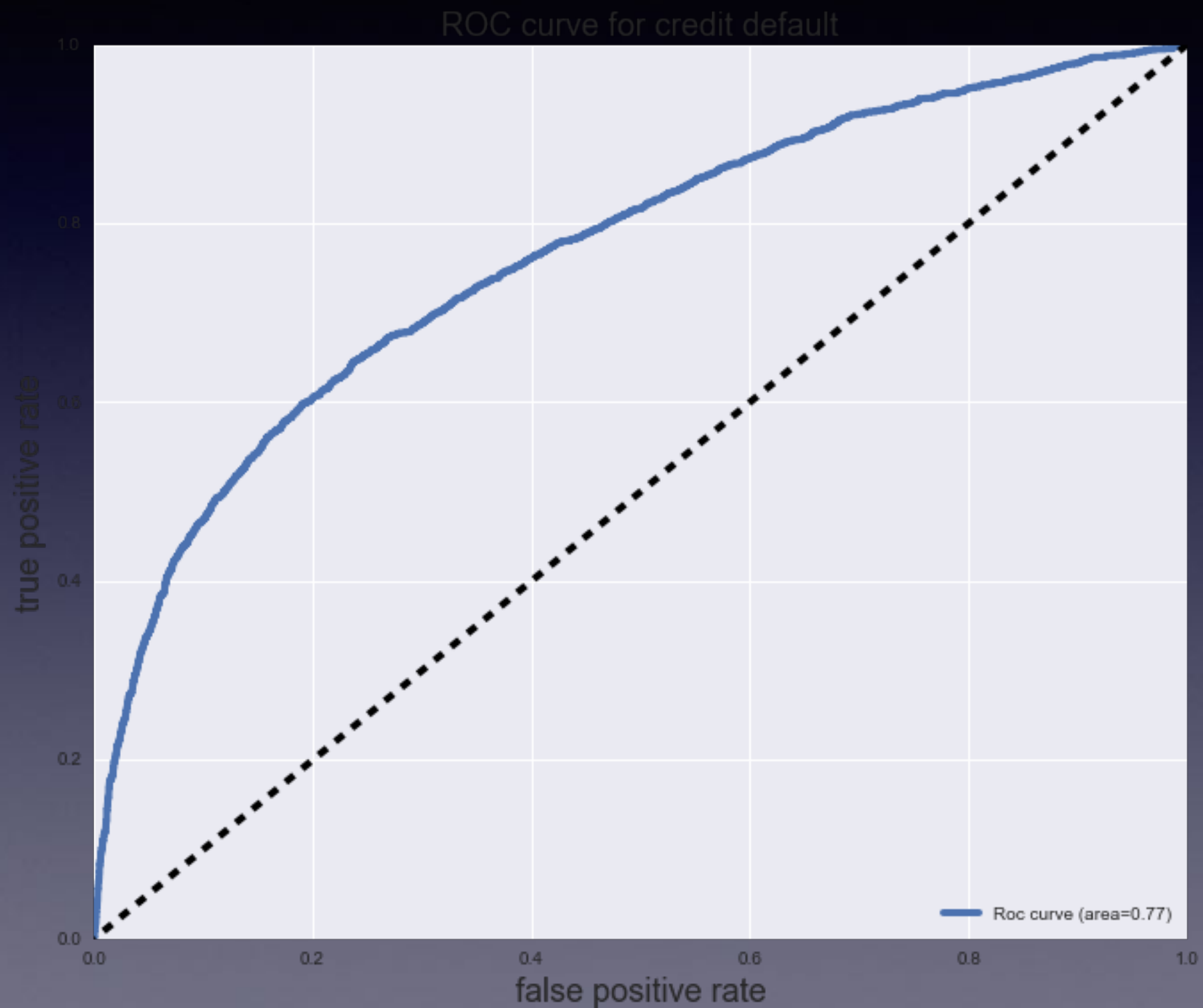
- Check correlation between variables to identify potential features
- Standardize numeric features
- Create dummy variables for categorical features
- Train_test_split with stratify

Models

- Logistic Regression with Gradient Descent (SGDClassifier) and Grid Search to optimize roc_auc
- Utilizing Lasso regularization to select features
- F1-score: 0.80, roc-auc: 0.77

	predicted default	predicted not default
default	731	1260
not default	380	6629

Models

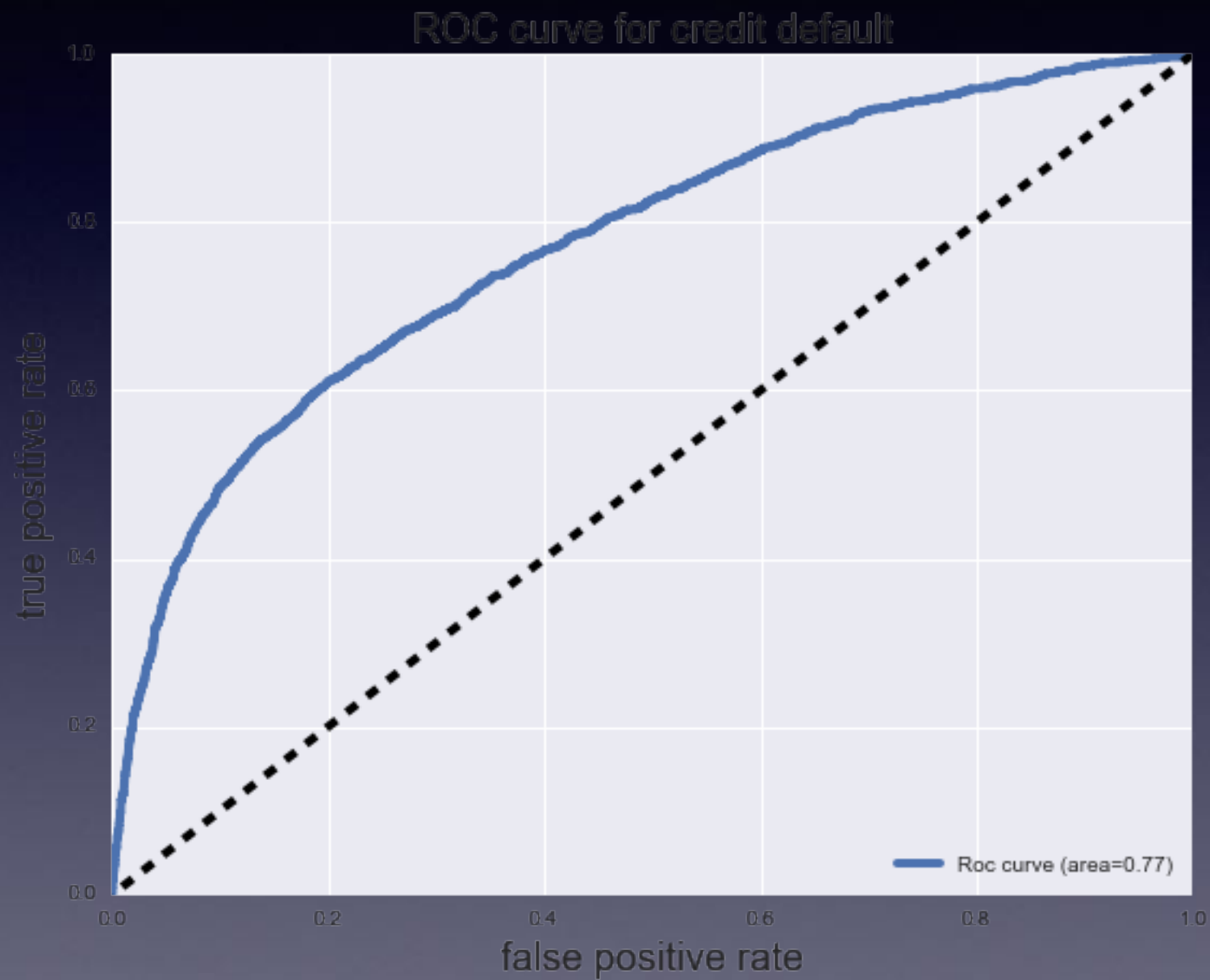


Models

- Logistic Regression with manual feature selection (RFECV) and Grid Search
- F1-score: 0.80, roc-auc: 0.77

	predicted default	predicted not default
default	708	1283
not default	344	6665

Models



Conclusion

- Potential model skewness due to unbalanced class
- Better dealing with outliers
- Potential multicollinearity