MUSIC GENRE CLASSIFICATION

KATE POLLEY

1.

BASELINE SYSTEM: LOGISTIC REGRESSION

LOGISTIC REGRESSION

- Trained 147 Logistic Regression classifiers
    - Liblinear solver
- GridSearchCV to learn C (regularization)
- Feature scaling with StandardScaler
- Train_test_split on training data
    - 6600 training, 1200 validation
- Cross validation score: 0.885
    - Higher than Kaggle score but close

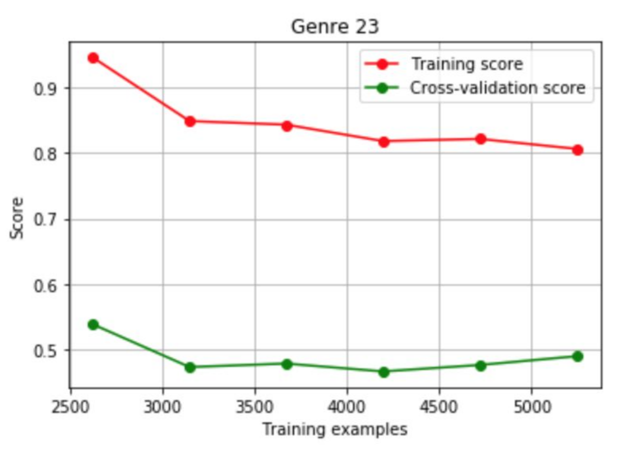
0.87565

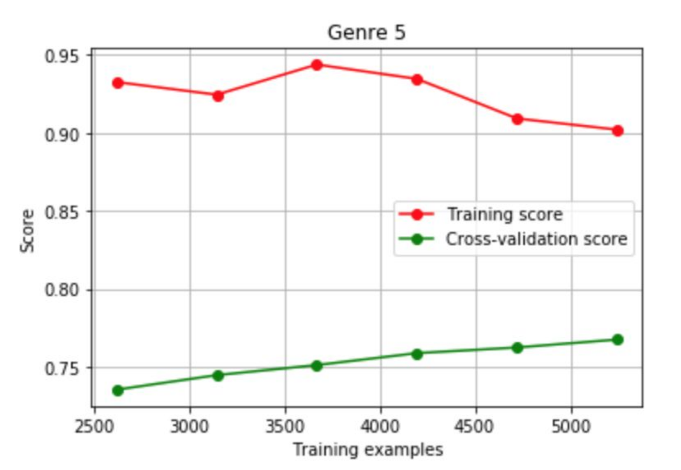
Best initial logistic regression score

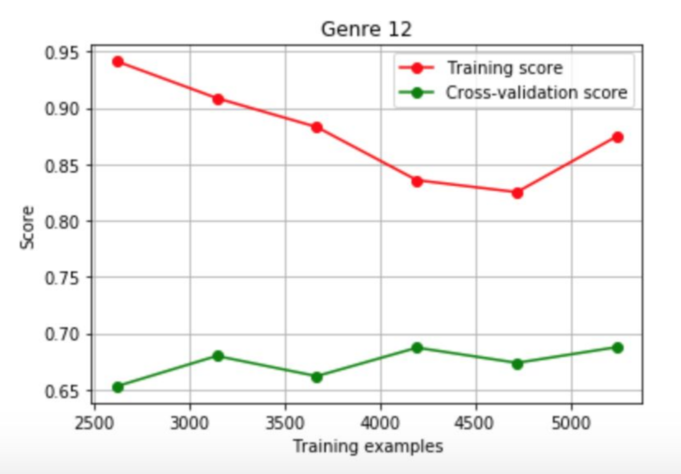
2.
EVALUATION OF CLASSIFIER PERFORMANCE

LEARNING CURVES

- Plotted learning curves for worst-performing classifiers
- Most revealed overfitting





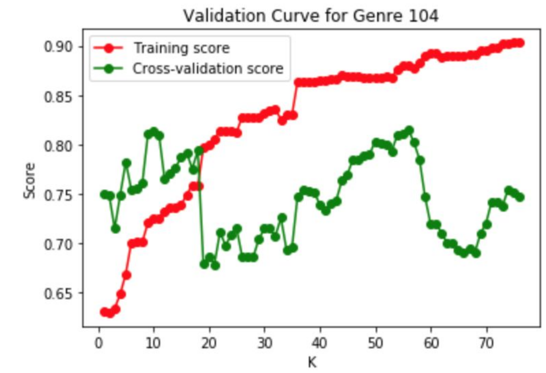


ADDRESSING OVERFITTING / HIGH VARIANCE

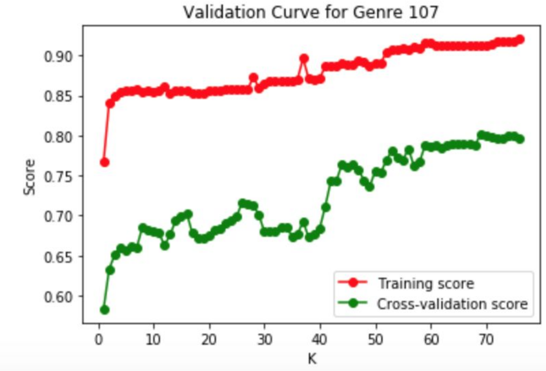
- More data - not exactly an option for this particular project
- Increase regularization - learned C parameters with strong regularization
- Fewer features - extract the most important features to train classifiers

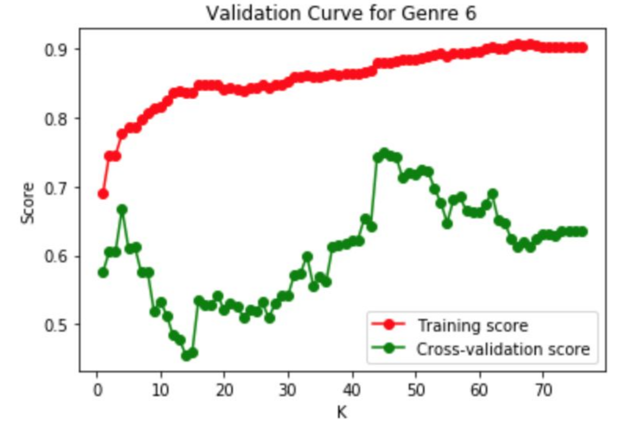
VALIDATION CURVES

- Determined if fewer features impacted cross validation score
- Used SelectKBest to select features


Validation Curve for Genre 104

Genre 107


Validation Curve for Genre 107


Validation Curve for Genre 6

LOGISTIC REGRESSION WITH FEATURE SELECTION

- Selected K by best performance on CV set
- Trained bad classifiers (<0.8 on initial test) with only K features
- Saw little improvement on CV score
  - CV score = 0.90391 (previously 0.885)
- Score on test data improved minimally
  - Issue: CV used to choose K and score

0.87626

Best score using feature selection

3.
USING OTHER KINDS OF CLASSIFIERS

NEURAL NETWORKS AND SVMS

- Only trained alternative classifiers where logistic regression performed below average
- Used grid search to learn parameters: hidden layer size and alpha / C and gamma
- Some performed much worse than previous classifiers, some performed slightly better
- Slow to run - not worth it

SELECTING CLASSIFIERS

- Combined best performing classifiers from original logistic regression, logistic regression with feature selection
- Neural networks and SVMs didn’t help
- Score on cross validation set increased
  - CV score = 0.9387 (previously 0.9039)
- Still failed to improve test score
  - Overlap with model selection + scoring

NEW VALIDATION SETUP

- Improve model selection for generalization
- Split into 3 sets: 1 training + 2 validation
  - ~ 5900, 1000, 1000
- Final CV score on classifiers: 0.8955
  - Closer to Kaggle score
- Train again on entire data set
  - Training score: 0.9387
  - Test score (Kaggle): 0.88112

0.88112

Best score with improved model selection

0.87565

Logistic regression, learned C

0.87626

Logistic regression, select K features

0.88112

Improved model selection

0.87553

Update: Final Leaderboard Score

Problem: inability to generalize

NEXT STEPS

- More in-depth evaluation of classifiers
  - Evaluation of latest models
  - Learning/validation curves
- Setup/Process
  - What worked, what didn’t
  - Areas for improvement
- Better address overfitting - Train > Test
- Feature design and selection

Fin