Optimize Feature Selection using Generic Algorithm/LASSO/Elastic Net

Hector Flores, Junlin Wang, Yi Lan

Abstract

- This study introduces optimization techniques for feature selection to improve classification accuracy for music genre classification and classification of whether a computer is a PC or Mac. We experimented with two well known optimization techniques: genetic algorithms and lasso methods for feature selection[1].
- We applied the techniques with a **Support Vector Machine (SVM)** classifier in order to address the music genre classification problem and PC/MAC classification problem with minimum features. We then compare top selected features and accuracy rate, to decide if it is a good selection.

Background

- **Feature selection** is the process of selecting a subset of the most relevant variables. These techniques can be used to make a simpler model for interpretation; shorter training times; avoid curse of dimensionality; reduce overfitting.
- **Least Absolute Shrinkage and Selection Operator (LASSO)** is commonly used for analyzing high-dimensional data. LASSO's objective is to find the parameter vector that minimizes the sum of squared errors and the L1 regularization term[2]:

$$\theta^* = \underset{\theta}{argmin}\left(\frac{1}{N}\|y - \boldsymbol{X}\theta\|_2^2 + \lambda\|\theta\|_1\right)$$

- **ElasticNet** extends LASSO by adding and additional L2 penalty term[3].
- **Genetic Algorithms (GA)** simulate natural evolution processes to find the solution for an optimization problem. Genetic algorithms seem to be an appropriate choice for selecting relevant features because they can deal with the search space generated by high number of features[4, 5].

Dataset

- This study used both audio and metadata provided by the **Free Music Archive** (FMA)[6]
  - The FMA dataset has 6400 instances, 518 features, and categorized into 8 classes[7]
  - All tracks are mp3-encoded, most of them with sampling rate of 44,100 Hz, bit rate 320 kbit/s (263 kbit/s on average), and in stereo.
  - The small balanced subset was used for lower computational resources
    - 8,000 30s clips from 8 top genres, balanced with 1,000 clips per genre, 1 root genre per clip.
- The **PCMAC** dataset is also used for testing.[8]
  - The PCMAC data set has 1943 instances, 3289 features, and categorized into 2 classes.

Methods

- A 80/10/10% split into training, validation, and test sets for the FMA data.
- A 70/30% split into training and test sets for the PC/MAC data.
- We trained an SVM classifier on the split FMA and PC/MAC dataset. We compared single top genre classification accuracy and number of features using:
  - No feature selection
  - LASSO
  - ElasticNet
  - Genetic Algorithm

Results

- Figures 1 and 4 displays the feature rankings for the FMA dataset and PC/MAC dataset, respectively.
- Figures 2 and 5 shows an increase in classification accuracy using feature selection for both datasets.
- Figures 3 and 6 shows the number of dimensions reduced for both datasets.

Figure 1. Ranking of feature importance for FMA dataset.

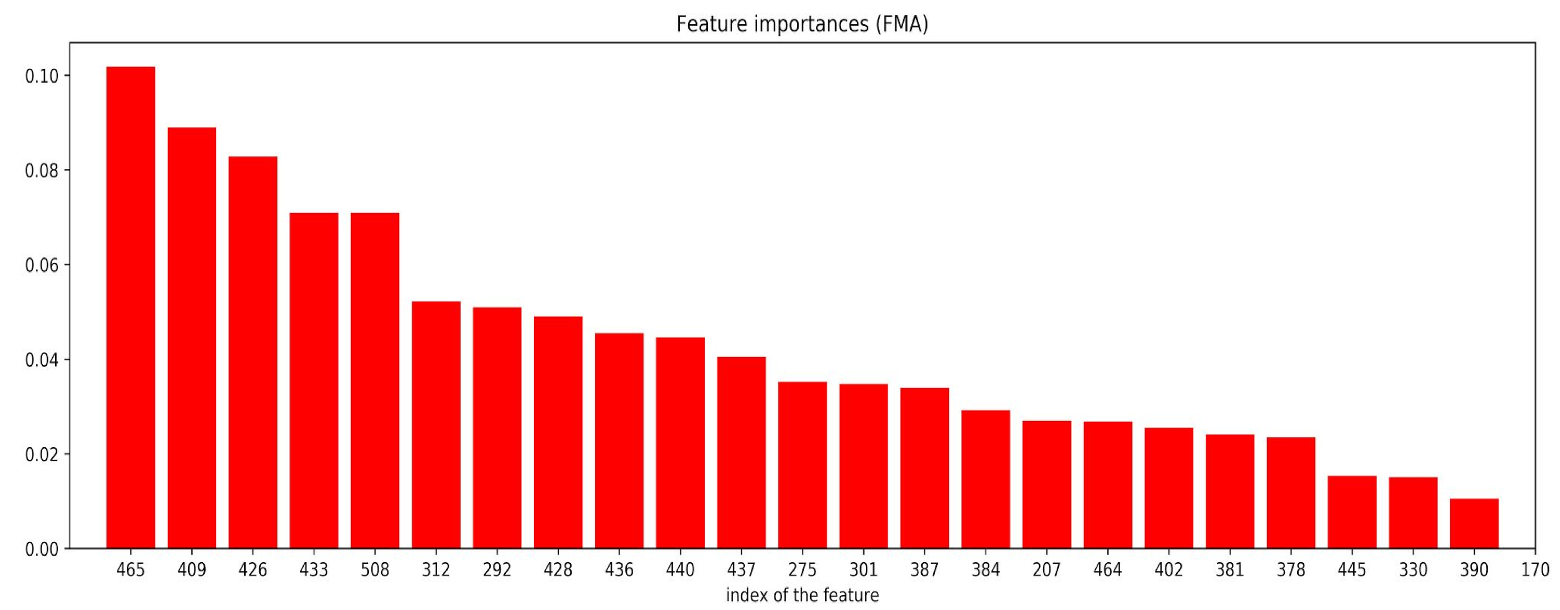


Figure 2. SVM classification accuracy for: all features, GA selection, and ElasticNet selection.

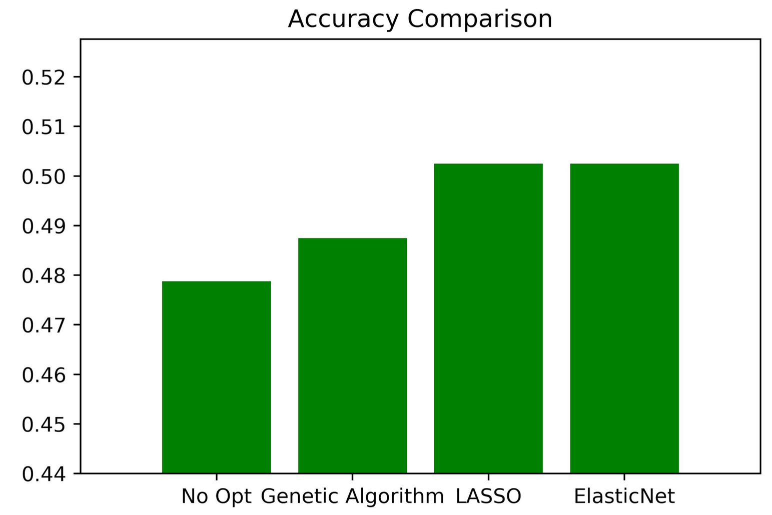


Figure 3. Number of features selected by: GA, LASSO, and ElasticNet.

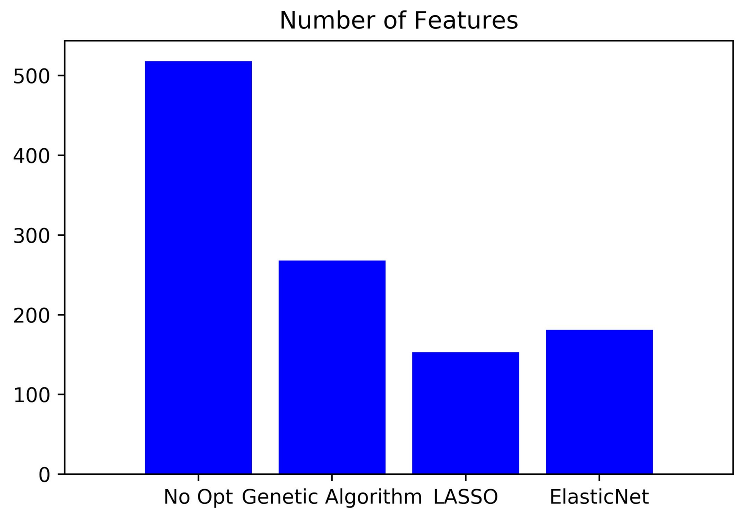


Figure 4. Ranking of feature importance for PCMAC dataset

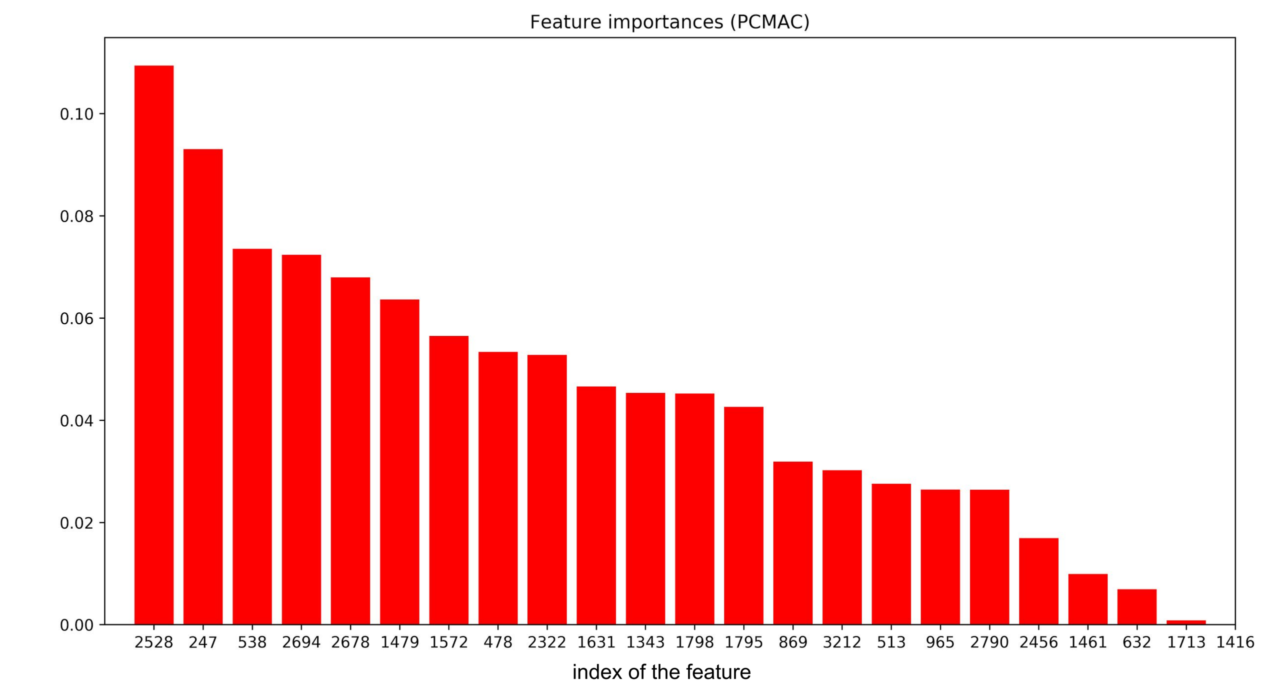


Figure 5. SVM classification accuracy for: all features, and ElasticNet selection.

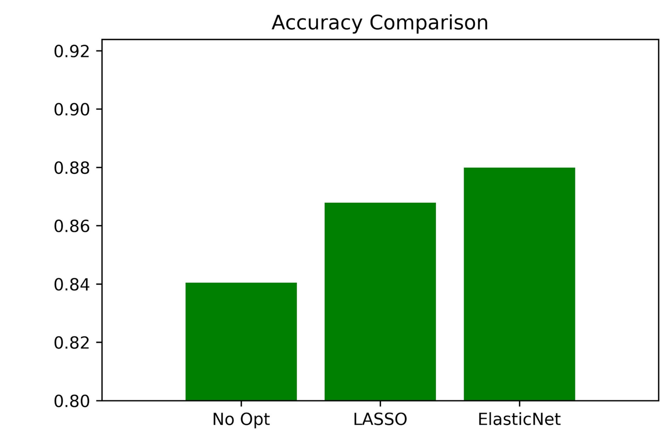
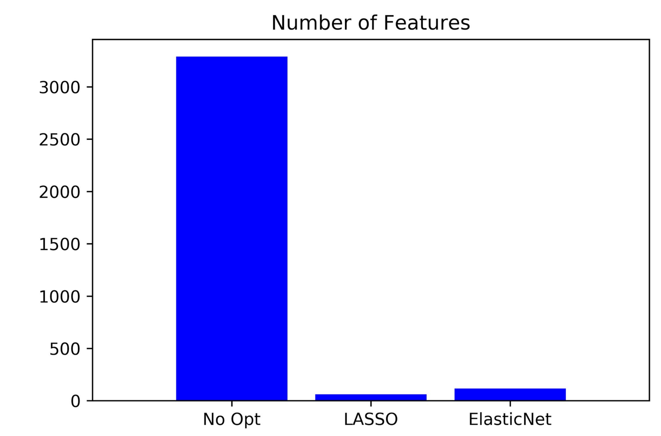


Figure 6. Number of features selected by: LASSO, and ElasticNet.



Conclusions

- LASSO and Elastic Net perform identically in terms of the increase in accuracy. Elastic Net achieves the same accuracy as LASSO in the FMA dataset and does marginally better than LASSO in the PCMAC dataset.
- However, LASSO always reduces the features by a larger amount. In the FMA dataset, LASSO reduces the number of features from 518 to 153, but Elastic Net reduces number of features to 181. In PCMAC dataset, LASSO reduces the number of features from 3289 to 62, whereas Elastic Net reduces it to 117.
- Genetic Algorithm takes a long time to train, and is outperformed by LASSO and Elastic Net. In the future we may look into Particle Swarm Optimization (PSO) which converges faster and may have a better performance.
- For all algorithms we used, we experienced a big reduction in number of features, and an increase of ~2% in terms of accuracy.

References

1. M. Dash and H. Liu, "Feature selection for classification," Intell. Data Anal., vol. 1, no. 1–4, pp. 131–156, 1997.
2. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.)267–288 (1996)
3. Zou, H. and Hastie, T. (2005), Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67: 301–320. doi:10.1111/j.1467-9868.2005.00503.x
4. de la Iglesia, B.: Evolutionary computation for feature selection in classification problems. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 3(6), 381–407 (2013). https://doi.org/10.1002/widm.1106
5. Xue, B., Zhang, M., Browne, W.N., Yao, X.: A survey on evolutionary computation approaches to feature selection. IEEE Trans. Evol. Comput. 20(4), 606–626 (2016). https://doi.org/10. 1109/TEVC.2015.2504420
6. http://freemusicarchive.org/
7. K. Benzi, M. Defferrard, P. Vandergheynst, and X. Bresson, "Fma: A dataset for music analysis," arXiv preprint arXiv:1612.01840, 2016.
8. "Feature Selection Datasets." Datasets | Feature Selection @ ASU, ASU, featureselection.asu.edu/datasets.php.