Polynomial Exclusive lasso-based k-nearest neighbor for analyzing high-dimensional data

Vuong Nguyen

Dr. Waterbury

Introduction

- This project focused on studying Exclusive Lasso-based KNN for high-dimensional data sets.
- **Goal:** predict the class of a data point y by finding the data points closest to it and considering their classes.
- The algorithm finds a sparse vector representing the observations that are relevant in predicting the class of new observations.
- For an input vector y, the class of y is predicted by considering the k largest coefficients of this sparse vector.

Exclusive Lasso

- The primary goal of the algorithm is to solve the optimizing problem:

$$\hat{\alpha}_E = \text{argmin}_\alpha \left\{ ||y - \mathbf{X}\alpha||_2^2 + \lambda \sum_{g=1}^{G} ||\alpha_g||_1^2 \right\} \quad (1)$$

  Note that the optimization problem in (1) imposes a linear relationship between the data points.
- The equation assumes that input data are grouped; that is, the n observations have been partitioned into G groups.
- The algorithm performs variable selection based on finding the optimal sparse coefficient vector $\hat{\alpha}_E$ that utilizes ℓ_1^2 norm exclusive lasso.
- This choice of norm ensures that the vector $\hat{\alpha}_E$ has intra-group sparsity (via the ℓ_1 norm) and inter-group non-sparsity (via the ℓ_2 norm).

Polynomial Exclusive Lasso (PEL)

- We sought to modify the algorithm so that it could more effectively classify data by levering latent higher-dimensional structure.
- As a motivating example, the optimization problem is changed into:

$$\tilde{\gamma} = \text{argmin}_\gamma \Bigg\{ ||y + y^2 - \mathbf{X}\alpha - \mathbf{X}^2\beta||_2^2 + \lambda \sum_{g=1}^{G} (||\alpha_g||_1^2 + ||\beta_g||_1^2) \Bigg\} \quad (2)$$

- The optimization problem in (2) introduces a coefficient vector β that represent the polynomial relationship between the predictors and the input data point y.

Data

- The algorithms (including some of the numerical optimization procedures) were implemented using Python. Data sets used in our study are data sets are benchmark data sets provided by UCI machine learning repository:
  1. Ionosphere: Multivariate data shows the classification of radar returns from the ionosphere. Contains 35 features and 351 observations.
  2. Sonar: multivariate data that shows the classification between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. Contains 60 features and 208 observations.
  3. Wine: the data shows the classification of wine using chemicals. The data set contains 13 features and 128 observations.
  4. LSVT Voice rehabilitation: The data contains 126 samples from 14 participants, with 309 features. The data represent binary classification of phonations considered 'acceptable' or 'unacceptable' based on voice rehabilitation treatment.
- We considered primarily high-dimensional data sets.

PEL Results

- The algorithm using (2) is given by:

  $\hat{\gamma}_E \leftarrow \hat{\alpha}_E + \hat{\beta}_E$
  $\tilde{\gamma} \leftarrow$ k largest elements of $\hat{\gamma}_E$
  $l_y \leftarrow \max_{c_j \in C} \sum_k^{i=1}$
  for $i = 1, ..., k$ **do**
    $d_{r,(y+y^2,x_i^{NN}+x_i^{2(NN)})} \leftarrow ||y+y^2-(\tilde{\alpha}_i x_i^{NN}+\tilde{\beta}_i x_i^{2(NN)})||_2$
  end for
  for $i = 1, ..., k$ **do**
    if $d_{\max}^{NN} \neq d_{\min}^{NN}$ **then**
      $w_i \leftarrow \frac{d_{\min}^{NN} - d_{r,(y+y^2,x_i^{NN}+x_i^{2(NN)})}}{d_{\max}^{NN} - d_{\min}^{NN}}$
    else
      $w_i \leftarrow 1$
    end if
  end for
  $l_y \leftarrow \max_{c_j \in C} \sum_{i=1}^{k} w_i \times \tilde{\gamma}_i^{c_j}$

- The optimization problem is solved using `scipy.optimize`'s SLSQP, k is chosen using `GridSearch CV`.
- The result shows that, while we achieved better accuracy scores classifying data for LSVT, other data sets saw a decrease in accuracy.

PEL with Another Embedding

- Observing the decline in accuracy score, we considered another optimization problem:

$$\hat{\alpha}_E = \text{argmin}_\alpha \left\{ ||\mathbf{Y}^2 - \mathbf{X}^2\alpha||_2^2 + \lambda \sum_{g=1}^{G} ||\alpha_g||_1^2 \right\} \quad (3)$$

- This approach outperformed the PEL algorithm using (2).

No.	Dataset	Accuracy of $\|\|\mathbf{Y}^2 - \mathbf{X}^2\alpha\|\|$
1	LSVT	57.89
2	vehicle	72.04
3	sonars	84.06
4	ionoshere	75.86
5	wine	93.22

The accuracy of the algorithm using the optimization problem in (3)

- With its improved accuracy, this algorithm, unlike the previous one, illustrates that leveraging the nonlinear associations in the data y can improve predictive accuracy.
- However, solving (3) repeatedly was computationally intensive (we repeatedly used Grid Search CV to find the optimal choices k and G).

Findings

We have shown that underlying characteristics of the data play an important roles in defining which optimization problems works well for classifying. Hence, we investigated an efficient optimization method for this type of Exclusive Lasso optimization problem. In the later part of the research project, we used Coordinate Descent to solve the primary optimization problem. However, we have not had enough to implement `Grid-Search` using this method. The accuracy when using method increased for the Wine and LSVT datasets but decreased slightly for the others.

Conclusions

- This work illustrates the nuances in choosing the correct optimization problem for high-dimensional classification problems.
- Choosing fitting methods for the problem requires a deep understanding of optimization problems numerical methods for solving them.
- This work introduced the idea of applying kernel functions (e.g., $x \mapsto x^2$) onto the optimization problem in the algorithm.
- This leaves us with questions regarding the correct selection of kernel functions; what are efficient methods for determining *how* the data should be embedded or transformed?

Future Research

- To further improve the idea of Exclusive Lasso-based KNN, we will focus on developing an effective optimization problem that generalizes to multiple datasets.
- This would also improve the run time of the algorithms overall since the run-time of solving optimization problem.
- Hence, we aim to move forward to implement more research in multi-kernel learning algorithm for our current research.

References and acknowledgements

- Qiu, L., Qu, Y., Shang, C., Yang, L., Chao, F., Shen, Q. (2021). Exclusive Lasso-based K-nearest-neighbor classification. Neural Computing and Applications, 33(21), 14247–14261. https://doi.org/10.1007/s00521-021-06069-5
- Campbell, F., Allen, G. I. (2017). Within group variable selection through the exclusive Lasso. Electronic Journal of Statistics, 11(2). https://doi.org/10.1214/17-ejs1317
- Welcome to the UC Irvine Machine Learning Repository. UCI Machine Learning Repository. (n.d.). https://archive.ics.uci.edu/
- We would like to thank the The Wiliam G. and Mary Ellen Bowen Endowed Fund for the generous support for this research project.