IE506: Machine Learning - Principles and Techniques **Jan-May 2025**

Final Review Report : Sparse Prediction with the k-Support Norm

Team Name: Neural Knights *Team Members: Kundan Pratap (24n0454), Sunrit Pal (24n0456)*

1 Problem Statement

The objective of our project was to implement and evaluate sparsity-inducing methods for prediction tasks, focusing on overcoming limitations of traditional techniques like Lasso, which can shrink too many coefficients to zero when features are highly correlated. We explored the k-support norm, a tighter and more precise convex relaxation of true sparsity combined with ℓ_2 penalty, aiming to improve generalization and predictive performance.

2 Work Done Before Stage-1 Review

Before the Stage-1 review, we focused on understanding the core theoretical aspects:

- Studied the basics of sparsity and challenges of enforcing it using the ℓ_0 norm (non-convex, NP-hard).
- Explored convex relaxations like the ℓ_1 norm and its limitations, especially when handling large coefficients.
- Understood the motivation for combining sparsity with an ℓ_2 constraint to improve sample complexity.
- Defined the combined constraint set $\{w : \|w\|_0 \leq k, \|w\|_2 \leq 1\}$ and derived the k-support norm.
- Compared the k-support norm with Elastic Net, analyzing tightness and theoretical benefits.

We focused on theoretical grounding and had not yet fully implemented the algorithms at this stage.

3 Comments/Inputs given during the Stage-1 Review

The following feedback was given during the Stage-1 review:

- Gain a clear understanding of Algorithm 1, particularly how the inequalities determine the correct values of l and r.
- Write a detailed mathematical derivation of the dual k-support norm.
- Implement both Algorithm 1 (Proximity Operator) and Algorithm 2 (Accelerated Gradient Descent) and test them on datasets.
- Apply the k-support norm to a novel setting (e.g., multiclass classification).

4 Addressing Comments after Stage-1 Review

We addressed the Stage-1 comments as follows:

- Developed a detailed understanding of Algorithm 1, especially the role of inequalities in selecting l and r.
- Completed the mathematical derivation of the dual k-support norm and validated the formulation.
- Fully implemented Algorithm 1 and Algorithm 2 and applied them to real datasets.
- For novelty, extended the k-support norm to multiclass classification tasks using the Iris and Yeast datasets.

5 Experiments and Replications of Algorithms in Paper

This section outlines the experimental settings and dataset details.

Synthetic Data

We followed the setup from [?] (Sec. 5, Example 4). The oracle vector:

$$w^* = (\underbrace{3, \ldots, 3}_{15}, \underbrace{0, \ldots, 0}_{25})$$

with $y = (w^*)^\top x + \mathcal{N}(0, 1)$.

Data generation:

- Components 1–5: same mean $Z_1 \sim \mathcal{N}(0, 1)$,
- Components 6–10: mean $Z_2 \sim \mathcal{N}(0, 1)$,
- Components 11–15: mean $Z_3 \sim \mathcal{N}(0, 1)$,
- Components 16–40: standard normal $\mathcal{N}(0, 1)$.

South African Heart Disease Dataset

The dataset contains 462 samples and 9 variables; the task is binary classification (presence/absence of coronary heart disease).

20 Newsgroups Dataset

This binary classification task uses a subset of the 20 Newsgroups dataset from the LIBSVM repository. Positive class: 10 groups (sci.*, comp.*, misc.forsale); negative class: other 10 groups.

Preprocessing:

- Removed words appearing in fewer than 3 documents.
- Data split: 14,000 train, 1000 validation, 4996 test samples.

6 Novel Settings and Experiments

We extended the k-support norm to multiclass classification.

Iris Dataset

Summary:

- 150 instances, 4 features: sepal length, sepal width, petal length, petal width.
- 3 classes: Iris-setosa, Iris-versicolor, Iris-virginica (50 samples each).

Yeast Dataset

Summary:

- 1484 samples,
- 8 numerical features + 1 class label,
- 10 classes (protein localization sites).

Results

Method	MSE (median)	MSE (SE)	Accuracy (median)	Accuracy (SE)
Lasso	0.181076	0.004876	0.71875	0.010644
Elastic Net	0.181135	0.004869	0.71875	0.011105
k-support	0.179434	0.005210	0.71875	0.010514

Table 1: South African Heart Disease: MSE and accuracy (50 repeats).

Method	Mean MSE (SE)
Lasso	0.9001 (0.0506)
Elastic Net	0.9215 (0.0508)
k-Support Norm	0.8551 (0.0425)

Table 2: Synthetic Data: Mean MSE (standard error in parentheses).

Method	Validation Accuracy	Test Accuracy
Lasso	76.10%	60.55%
Elastic Net	76.30%	61.27%
k-Support Norm	78.20%	69.48%

Table 3: 20 Newsgroups: Validation and Test Accuracy.

Method	Accuracy
Lasso	0.9667
Elastic Net	0.9667
k-Support Norm	0.9667

Table 4: Iris Dataset: Classification Accuracy.

Method	Accuracy
Lasso	0.6296
Elastic Net	0.6195
k-Support Norm	0.6397

Table 5: Yeast Dataset: Classification Accuracy.

7 Conclusion

We successfully understood the k-support norm, including its mathematical formulation and advantages over Lasso and Elastic Net. We gained a thorough understanding of Algorithms 1 and 2 used to compute the k-support norm, with detailed insight into how the inequalities in Algorithm 1 determine the parameters l and r. We implemented the algorithms on the same three datasets as described in the research paper and obtained results that demonstrate the superior predictive performance of the k-support norm. Additionally, we extended the use of the k-support norm beyond the original scope by applying it to multiclass classification problems, specifically the Iris and Yeast datasets, and visualized the resulting decision boundaries.

8 Contributions

Both team members collaborated on studying the research paper and preparing the slides.

Sunrit Pal:

- Understood and explained the detailed derivation of the dual norm of the k-support norm as per the Stage-1 feedback.
- Provided an in-depth understanding of Algorithms 1 and 2, especially clarifying how the two inequalities in Algorithm 1 guide the selection of l and r.

Kundan Pratap:

- Implemented the code for both classification and regression algorithms and applied them to the datasets used in the paper.
- Developed and implemented the code for the novelty task (multiclass classification) and tested it on the Iris and Yeast datasets.

9 References

- Kenta Nakai, Paul Horton. *Yeast Dataset.* Available at: `https://www.kaggle.com/datasets/samanemami/yeastcsv`
- *Understanding Regularization Techniques: Lasso vs Ridge vs Elastic Net.* GeeksforGeeks. Available at: `https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/`
- *Understanding the Dual of Optimization Problems.* University of Washington Math Notes. Available at: `https://sites.math.washington.edu/~burke/crs/407/notes/section4.pdf`