# Reproducing Sequential Attention for Feature Selection

Gary Wang, Hank Lai
N16132087, N16131722
n16132087@gs.ncku.edu.tw, n16131722@gs.ncku.edu.tw

June 15, 2025

## 1 Introduction

Feature selection is essential for improving interpretability, generalization, and efficiency in high-dimensional learning tasks. However, As noted in a review by Nordling (2013, Chapter 5.2) have noted, it remains a challenging problem due to computational limits and sensitivity to noise. Classical methods such as LASSO and greedy search either fail to capture feature interactions or lack robustness.

Sequential Attention (SA) [1] proposes a novel attention-based relaxation of greedy selection that models conditional importance adaptively. In this study, we aim to reproduce SA and evaluate whether it consistently outperforms classical methods like Sequential LASSO and Group LASSO.

## 2 Problem Description

Given a dataset $X \in \mathbb{R}^{n \times d}$ and labels $y \in \mathbb{R}^n$, the goal of feature selection is to find a subset $S \subseteq \{1, ..., d\}$ of $k$ features that maximize the predictive performance of a model $f(X_S)$.

Classical methods often ignore the marginal gain of features conditioned on already-selected features. SA addresses this by learning attention scores in a sequential manner. We evaluate whether SA more effectively identifies relevant subsets by comparing its classification accuracy and stability to that of Sequential LASSO and Group LASSO on benchmark datasets.

## 3 Methods

### 3.1 Sequential Attention Algorithm

### 3.2 Experimental Setup

## 4 Results

### 4.1 Dataset Overview

### 4.2 Feature Selection Accuracy

### 4.3 Statistical Validation

## 5 Discussion

## 6 Conclusion

## References

[1] Taisuke Yasuda, MohammadHossein Bateni, et al. Sequential attention for feature selection. *International Conference on Learning Representations (ICLR)*, 2023.

# Appendix

## A Hyperparameters

## B Additional Figures