

Homework 2: CART, Logistic, LASSO

Students: Harsh Chaturvedi, Pratik Anand

Problem 1

Part a

A CART estimator trained with standardized data $\tilde{\mathbf{x}}$ will be identical to an estimator trained with un-standardized data \mathbf{x} . Running both estimators on unseen test data will produce identical results.

The reasoning for this is as follows.

- The CART estimator works by examining the data points over one dimension at a time, and determining the optimal point along that dimension where it should make a cut.
- Standardization changes the mean and variance of the data points along each dimension. It does not however change the ordering of the points or even the ratios of the inter-point distances in any dimension.
- The loss function (RMSE, GINI, cross-entropy etc) used to determine the cuts in the feature space therefore continue to attain minimum at the same relative position among the data points, along the dimension under consideration, regardless of standardization.

Part b

A LASSO regression estimator trained with standardized data $\tilde{\mathbf{x}}$ may be different from an estimator trained with un-standardized data \mathbf{x} . Running both estimators on unseen test data may produce different results.

The reasoning for this is as follows.

- The LASSO regressor works by adding a regularization term proportional to the sum of the magnitudes of the weights β_k 's to the loss function to be minimized. This regularization term is given by $\lambda \sum_k^p |\beta_k|$.
- The weight β_k for a feature / dimension k will be significantly different depending on whether the data has been standardized along that dimension or not. Thus the relative proportions of the different $|\beta_k|$'s in the regularization term will differ greatly for standardized and un-standardized data.
- As a result, equal changes in different β_k 's will produce comparable changes in the regularization term if the data are standardized, while equal changes in different β_k 's will have very different magnitudes of impact on the regularization term if the data are not standardized.
- Thus, standardization will result in a different (usually better) set of optimal weights β_k 's than a model without standardization.

Problem 2

We wish to solve the following minimization problem

$$\min_{\beta} \|Y - X\beta\|^2 \text{ subject to } \sum_k^p |\beta_k| \leq s$$

The objective is to minimize a function of one variable subject to one constraint. The Lagrangian for this problem will be given by,

$$L(\beta, \lambda) = \|Y - X\beta\|^2 + \lambda(\sum_k^p |\beta_k| - s) \text{ where } \lambda \text{ is a Lagrange multiplier.}$$

And the solution to our problem will be obtained by solving,

$$\nabla_{\beta} L(\beta, \lambda) = 0$$

$$\implies \nabla_{\beta} \|Y - X\beta\|^2 + \lambda \nabla_{\beta} \sum_k^p |\beta_k| - \nabla_{\beta}(s) = 0. \text{ Since } s \text{ is a constant,}$$

$$\implies \nabla_{\beta} (\|Y - X\beta\|^2 + \lambda \sum_k^p |\beta_k|) = 0$$

The solutions of this equation represent the extrema and inflection points of the bracketed expression. The minimum of the expression, assuming one exists, is therefore a solution,

$$\min_{\beta} (\|Y - X\beta\|^2 + \lambda \sum_k^p |\beta_k|)$$

Note: By this proof, the β_k^c 's are equivalent to the β_k 's as defined in the original problem statement.

In []: