

# inference

*Wesley Brooks*

## Introduction

It has often been noted that the estimation error underestimates the prediction error over novel data generated from the same process. Stein's unbiased risk estimation (SURE) is a framework for estimating the prediction error of a model. The so-called covariance penalty estimate arises from the SURE framework as an estimate of the degrees of freedom used in estimating the model.

Equivalent computations arise from the framework of Akaike's information criterion (AIC). The AIC arises from the framework of estimating the likelihood of a fitted model with respect to the unknowable truth that gave rise to the data.

## Methods

Assume that we observe data  $\mathbf{y} = (y_1, \dots, y_n)$  such that  $y_i = m(\mu_i)$  where  $\mu_i = Ey_i$ .

By the covariance penalty, the degrees of freedom used in estimating a model is given by

$$df = \sum_{i=1}^n cov(y_i, \mu_i) = \sum_{i=1}^n \frac{d}{d\mu_i} m(\mu_i).$$

But it has been noted that the degrees of freedom is also given by the number of covariates with nonzero coefficients.

The divergence formula of [Zou-Hastie-Tibshirani-2007] states that

$$\left( \frac{\partial \hat{\boldsymbol{\mu}}}{\partial \mathbf{y}} \right)_{i,j} = \frac{\partial \hat{\mu}_i}{\partial y_j}, i, j = 1, 2, \dots, n.$$

Thus, the divergence

$$\nabla \cdot \hat{\boldsymbol{\mu}} = \text{tr} \left( \frac{\partial \hat{\boldsymbol{\mu}}}{\partial \mathbf{y}} \right).$$

Now in the case of a model estimated by LAGR, there is an independent model for each observation, so the derivative must be calculated independently for each:

$$\text{tr} \left( \frac{\partial \hat{\boldsymbol{\mu}}}{\partial \mathbf{y}} \right) = \sum_{i=1}^n \frac{\partial \hat{\mu}_i}{\partial y_i}$$

where

$$\frac{\partial \hat{\mu}_i}{\partial y_i} = ?$$