算法介绍

吴秉哲

2015 年 9 月 30 日

1 Algorithm Introduction

In this project,we used five algorithms.This is the introduction to these algorithms.

1.1 The Linear Regression

Linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables denoted X.In this project ,linear regression fits a model with coefficents$\omega = (\omega_1, \omega_2, \cdots, \omega_p)$ to minimize the residual sum of squares between the observed responses in the dataset,and the responses predicted by the linear approximation.Mathematically,it solves a problem of the term:

$$\min_{\omega} ||X\omega - y||_2{}^2 \tag{1}$$

1.2 Ridge Regression

Ridge regression addresses some of the problems of linear regression by imposing a penalty on the size of coefficients.The ridge coefficients minmize a penalized residual sum of squares,

$$\min_{\omega} ||X\omega - y||_2{}^2 + \alpha||\omega||_2^2$$

Here,$\alpha \geq 0$is a complexity parameter that controls the amount of shrinkage:the larger value of α the greater the amount of shrinkage and thus the coefficients become more robust to collinearity.

1.3 Lasso

The Lasso is a linear model that the estimates sparse coefficients.It is useful in some contexts due to its tendency to prefer solutions with fewer parameter

values ,effectively reducing the number of variables upon which the given solution is dependent.For this reason,the Lasso its variants are fundamental to the field of compressed sensing . Under certain conditions,it can recover the exact set of non-zero weights.

Mathematically,it consists of a linear model trained with $\updownarrow_1$ prior as regularizer.The objective function to minimize is:

$$\min_{\omega} \frac{1}{2n_{samples}}||X\omega - y||_2^2 + \alpha||w||_1$$

The Lasso estimate thus solves the minimization of the least-squares penalty with $\alpha||\omega||_1$ added, where $\alpha$is a constant and $||\omega||_1$is the $\updownarrow_1$-norm of the parameter vector.

1.4 Gradient Tree Boosting

Gradient Boosted Regression Trees(GBRT) is a generalizition of boosting to arbitary differentiable loss function.GBRT is an accurate and effective procedure that can be used for both regression and classification problems.