Algorithm Introduction

WuBingzhe

2015 年 10 月 2 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:

$$\underset{\omega}{min} \, ||X\omega - y||_2{}^2 \tag{1}$$

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,

$$\underset{\omega}{min} \, ||X\omega - y||_2{}^2 + \alpha||\omega||_2^2$$

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:

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

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.Like other boosting methods,gradients boosting commbines weak learners into a single strong learner,in an iterative fashion.It is easiest to explain in the least-squares regression setting,where the goal is to learn a model F that predicts values$\hat{y} = F(x)$,minimizing the mean mean squared error$(\hat{y} - y)^2$to the true values y.

GBRT considers additive models of the following form:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x)$$

where$h_m(x)$are the basis functions which are usually called weak learners in the context of boosting. Gradient Tree Boosting uses decision trees of fixed of size as weak learns.Decision trees have a number of abilities that make them valuable for boosting algorithms GBRT builds the additive model in a forward stagewise fashion:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

At each stage the decison tree $h_m(x)$ is chosen to minimize the loss function L given the current model F_{m-1} and its fit $F_{m-1}(x_i)$:

$$F_m(x) = F_{m-1}(x) + \underset{h}{argmin} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) - h(x))$$

The initial modelF_0 is problem specific,for least-squares regression one usually chooses the mean of the target values.

Gradient Boosting attempts to solve this minimization problem numerically via steepest descent direction is the negative gradient of the loss function evaluated at the current model F_{m-1} which can becalculated for any differentiable loss function:

$$F_m(x) = F_{m-1}(x) + \gamma_m \sum_{i=1}^{n} \nabla_F L(y_i, F_{m-1}(x_i))$$

Where the step length γ_m is chosen in line searh:

$$\gamma_m = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) - \gamma \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)})$$