

Smoothing Splines

Vito Simone Lacatena mat.747810

3/7/2021

1. Introduzione

Il modello di regressione polinomiale estende il modello di regressione lineare attraverso l'aggiunta di ulteriori variabili ottenuti dalla variabile originaria elevandoli ad una potenza. Ovvero si estende il modello lineare classico

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

con la seguente funzione polinomiale:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_d x_i^d + \epsilon_i$$

2. Struttura dell'algoritmo

2.1 Concetto di Spline

Una Spline è una funzione continua costituita da un insieme di polinomi raccordati tra loro, il cui scopo è interpolare in un intervallo un insieme di punti (detti nodi della spline), in modo tale che la funzione sia continua almeno fino ad un dato ordine di derivate in ogni punto dell'intervallo.

2.2 Smoothing Splines

L'obiettivo della regressione è adattare una funzione $f(x)$ ad un set di dati tale che minimizzi la somma dei quadrati dei residui

$$RSS = \sum_{i=1}^n (y_i - f(x_i))^2$$

In questo modo però a quale problema si andrebbe incontro? Si potrebbe ottenere un RSS pari a zero scegliendo la funzione f che interpoli perfettamente tutti i punti y_i , ma tale risultato sarebbe troppo soggetto ad overfitting poiché la funzione sarebbe troppo flessibile. Occorre quindi trovare il modo di smussare la funzione, ridefinendo la funzione RSS nel seguente modo:

$$RSS(f, \lambda) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int f''(t)^2 dt$$

La funzione f è definita *smoothing spline*.

La prima parte di RSS definisce una misura di distanza dai dati mentre la seconda parte definisce una penalità che penalizza la variabilità della funzione f

2.3

Descrivere in dettaglio la struttura dell'algoritmo, anche nel caso multidimensionale;

3. Implementazioni

Illustrare le principali implementazioni disponibili in R;

4. Esempio con dati reali

presentare un esempio realistico in R.