HỒI QUY TUYẾN TÍNH

Linear Regression

TS. Nguyễn Thị Kim Ngân

Giới thiệu

Bài toán:

Cho: 1000 căn nhà trong thành phố, mỗi căn nhà:

rộng x_1 m2,

có x_2 phòng ngủ

cách trung tâm thành phố x_3 km

giá của mỗi căn nhà

Hỏi: giá của căn nhà thứ 1001 là bao nhiêu?

Hàm dự đoán y = f(x) có dạng thế nào?

Giới thiệu

Nhận xét:

i) Diện tích nhà càng lớn thì giá nhà càng cao

ii) Số lượng phòng ngủ càng lớn thì giá nhà càng cao

iii) Càng xa trung tâm thì giá nhà càng giảm.

=>Dựa trên quan sát này, ta có thể mô hình quan hệ giữa đầu ra và đầu vào bằng một hàm tuyến tính đơn giản

Giới thiệu

Vector đặc trưng x = $[x_1, x_2, x_3]$

Vector trọng số (weight vector) w = $[w_1, w_2, w_3]$

Hàm số f(x) mô tả mối quan hệ giữa giá nhà và 3 đại lượng đầu vào:

$$f(\mathrm{x}) = w_1\mathrm{x}_1 + w_2\mathrm{x}_2 + w_3\mathrm{x}_3$$

Giới thiệu

$$f(\mathrm{x}) = w_1\mathrm{x}_1 + w_2\mathrm{x}_2 + w_3\mathrm{x}_3$$

$$\mathrm{y} \approx f(\mathrm{x}) = \hat{\mathrm{y}}$$

Mối quan hệ y≈f(x) là một mối quan hệ tuyến tính (linear)

Bài toán trên là bài toán thuộc loại regression.

Do đó, bài toán đi tìm các hệ số tối ưu $\{w_1, w_2, w_3\}$ được gọi là bài toán Linear Regression

Giới thiệu

$$f(\mathrm{x}) = w_1\mathrm{x}_1 + w_2\mathrm{x}_2 + w_3\mathrm{x}_3$$

$$\mathrm{y} \approx f(\mathrm{x}) = \hat{\mathrm{y}}$$

i) y và ŷ là hai giá trị khác nhau do có sai số mô hình, chúng ta mong muốn rằng sự khác nhau này rất nhỏ

ii) Linear là thẳng, phẳng

- Trong không gian hai chiều, đồ thị của một hàm số tuyến tính là một đường thẳng
- Trong không gian ba chiều, đồ thị của một hàm số tuyến tính là một mặt phẳng
- Trong không gian nhiều hơn 3 chiều, đồ thị của một hàm số tuyến tính là một *siêu phẳng* (*hyperplane*)

Dạng của Linear Regression

Vector đặc trưng x= $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_m \end{bmatrix}$ => $x^T = [x_1, x_2, x_3, \dots, x_m]$

Vector trọng số (weight vector) w = $\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \\ w_m \end{bmatrix}$

Hàm số f(x) mô tả mối quan hệ giữa các đại lượng đầu vào:

$$f(\mathrm{x}) = w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + w_m x_m$$

$$\hat{\mathrm{y}} = \mathrm{f(x)} = \mathrm{x}^T \mathrm{w}$$

$$\mathrm{y} \approx \hat{\mathrm{y}} = \mathrm{x}^T \mathrm{w}$$

Sai số dự đoán

$$y \approx \hat{y} = x^T w$$

Chúng ta muốn sự sai khác e giữa giá trị thực y và giá trị dự đoán $\hat{y}$ là nhỏ nhất:

$$\frac{1}{2}e^2 = \frac{1}{2}(y - \hat{y})^2 = \frac{1}{2}(y - x^T w)^2$$

- Hệ số $\frac{1}{2}$ là để thuận tiện cho việc tính toán (khi tính đạo hàm thì số $\frac{1}{2}$ sẽ bị triệt tiêu)
- Sự sai khác giữa y và $\hat{y}$ không đồng nghĩa với e nhỏ nhất. Vì $e = y - \hat{y}$ có thể là một số âm, khi $e = -\infty$ là rất nhỏ nhưng sự sai lệch là rất lớn. Do đó chúng ta cần e^2

Hàm mất mát

Với tất cả các cặp dữ liệu quan sát được (x_i , y_i), i = 1, 2, . . . , N, chúng ta muốn trung bình sai số là nhỏ nhất

=>Tìm w để hàm số sau đạt giá trị nhỏ nhất:

$$\mathcal{L}(\mathrm{w}) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - x_i^T w)^2$$

=> Cần tối thiểu hàm mất mát $\mathcal{L}(w)$ theo w. Nghiệm cần tìm của bài toán là:

$$w^* = \underset{w}{\operatorname{argmin}}\, \mathcal{L}(w)$$

Hàm mất mát

Với tất cả các cặp dữ liệu quan sát được (x_i , y_i), i = 1, 2, . . . , N, chúng ta muốn $\mathcal{L}(w)$ nhỏ nhất

$$\mathcal{L}(\mathrm{w}) = \frac{1}{2\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \mathrm{x_i^T w})^2$$

Đặt $y = [y_1, y_2, \ldots, y_N]^T$, $X = [x_1, x_2, \ldots, x_N]$, $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_m \end{bmatrix}$

Hàm $\mathcal{L}(w)$ được viết lại như sau

$$\mathcal{L}(\mathrm{w}) = \frac{1}{2\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \mathrm{x_i^T w})^2 = \frac{1}{2\mathrm{N}}\left\|\begin{bmatrix} \mathrm{y_1} \\ \mathrm{y_2} \\ \cdots \\ \mathrm{y_N} \end{bmatrix} - \begin{bmatrix} \mathrm{x_1^T} \\ \mathrm{x_2^T} \\ \cdots \\ \mathrm{x_N^T} \end{bmatrix}\mathrm{w}\right\|^2 = \frac{1}{2\mathrm{N}}\|\mathrm{y} - \mathrm{X^T w}\|_2^2$$

Nghiệm cho bài toán Linear Regression

$$\mathcal{L}(w) = \frac{1}{2N}\|y - X^T w\|_2^2$$

Nhận thấy rằng hàm mất mát $\mathcal{L}(w)$ có đạo hàm tại mọi w (xem Bảng 2.1)

=> Tìm giá trị tối ưu của w có thể được thực hiện thông qua việc giải phương trình đạo hàm của $\mathcal{L}(w)$ theo w bằng không

Đạo hàm theo w của hàm $\mathcal{L}(w)$ là:

$$\frac{\nabla \mathcal{L}(\mathbf{w})}{\nabla \mathbf{w}} = \frac{1}{N}\mathbf{X}(\mathbf{X}^T\mathbf{w} - \mathbf{y})$$

Giải phương trình đạo hàm bằng không:

$$\frac{\nabla \mathcal{L}(\mathbf{w})}{\nabla \mathbf{w}} = 0 \Leftrightarrow \mathbf{X}\mathbf{X}^T\mathbf{w} = \mathbf{X}\mathbf{y}$$

Nghiệm cho bài toán Linear Regression

Giải phương trình đạo hàm bằng không:

$$\frac{\nabla \mathcal{L}(\mathbf{w})}{\nabla \mathbf{w}} = \mathbf{0} \Leftrightarrow \mathbf{X}\mathbf{X}^T\mathbf{w} = \mathbf{X}\mathbf{y}$$

- Nếu ma trận XX^T khả nghịch thì phương trình trên có nghiệm duy nhất là

$$\mathbf{w} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{y}$$

- Nếu ma trận XX^T không khả nghịch thì nghiệm của phương trình có thể xác định dựa vào giả nghịch đảo

$$\mathbf{w} = (\mathbf{X}\mathbf{X}^T)^{\dagger}\mathbf{X}\mathbf{y}$$

Ví dụ

ID nhà	Diện tích (x_1)	Số phòng ngủ (x_2)	Cách trung tâm (x_3)	Giá
1	60	2	10	10
2	40	2	5	12
3	100	3	7	20

Dự đoán giá của căn nhà x=(50, 2, 8)?
Xây dựng tập các vector đầu vào (X), và vector giá y
Tính tham số w của hàm f(x).

$$\mathbf{w} = (\mathbf{X}\mathbf{X}^T)^\dagger \mathbf{X}\mathbf{y}$$

Bài toán

Chúng ta có 1 bảng dữ liệu về chiều cao và cân nặng của 15 người:

STT	Chiều cao (cm)	Cân nặng (kg)
1	147	49
2	150	50
3	153	51
4	155	52
5	158	54
6	160	56
7	163	58
8	165	59

STT	Chiều cao (cm)	Cân nặng (kg)
9	168	60
10	170	72
11	173	63
12	175	64
13	178	66
14	180	67
15	183	68

Bài toán

Bài toán đặt ra là: dự đoán cân nặng của một người dựa vào chiều cao của họ

Ta nhận thấy: cân nặng tỉ lệ thuận với chiều cao (càng cao càng nặng)

Để kiểm tra độ chính xác của model tìm được, ta sẽ giữ lại dòng 4 và 6 để kiểm thử

w = [[-33.73541021]
[0.55920496]]

y1 = w_1*155 + w_0

y2 = w_1*160 + w_0

Bias trick

Thêm hệ số b vào sẽ khiến cho mô hình linh hoạt hơn một chút bằng cách bỏ ràng buộc đường thẳng quan hệ giữa đầu ra và đầu vào luôn đi qua gốc toạ độ

$$y = x^T w + b$$

Đặt $x_0=1$

$$\bar{\mathbf{x}} = [x_0, x_1, x_2, \ldots, x_N]^T$$
$$\bar{\mathbf{w}} = [b, w_1, w_2, \ldots, w_N]$$
$$\bar{\mathbf{X}} = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_N]$$

Có phương trình

$$y = \mathbf{x}^T\mathbf{w} + b = w_1x_1 + w_2x_2 + \cdots + w_dx_d + bx_0 = \bar{\mathbf{x}}^T\bar{\mathbf{w}}$$

Bias trick

Có phương trình

$$y = \mathbf{x}^T\mathbf{w} + b = w_1x_1 + w_2x_2 + \cdots + w_dx_d + bx_0 = \bar{\mathbf{x}}^T\bar{\mathbf{w}}$$

Nghiệm

$$\bar{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2N}\|\mathbf{y} - \bar{\mathbf{X}}^T\bar{\mathbf{w}}\|_2^2 = (\bar{\mathbf{X}}\bar{\mathbf{X}}^T)^\dagger\bar{\mathbf{X}}\mathbf{y}$$