Selecting meaningful features

Nhóm 4

DamSanX

7/2020

Mục lục

Mục lục

Introduction

Trong thực tế dữ liệu thu thập của chúng ta có rất nhiều feature, nhưng không phải feature nào cũng có ý nghĩa cho việc xây dựng mô hình dự đoán. Một ví dụ điển hình đó là việc chẩn đoán bệnh đau lưng.

$\Rightarrow$ Selecting meaningful features

Mục lục

L1 and L2 regularization as penalties against model complexity

Như đã được giới thiệu ở chương 3, **L2 regularizaton** là một hướng cách để giảm độ phức tạp của mô hình bằng việc trừng phạt những w_i có giá trị lớn.
Ta định nghĩa chuẩn L2 của vector w:

$$L2: \qquad \|w\|_2^2 = \sum_{j=1}^{m} w_j^2$$

L1 and L2 regularization as penalties against model complexity

Một hướng tiếp cận khác để giảm sự phức tạp của mô hình đó là sử dụng **L1 regularization**:

$$L1: \qquad \|w\|_1 = \sum_{j=1}^{m} |w_j|$$

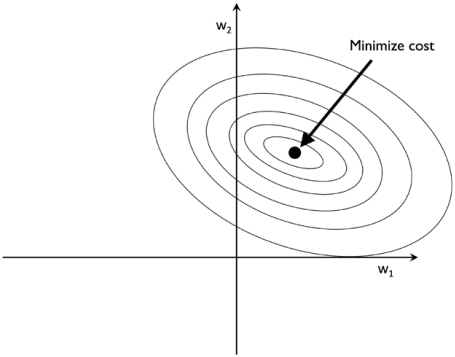
L1 regularization có thể làm thưa vector w (sparsity), nên có thể hiểu đây là một kỹ thuật cho feature selecion.

Mục lục

A geometric interpretation of L2 regularization

Chúng ta lấy ví dụ về cost function được dùng trong Adaline, đó là **SSE** (sum of squared errors):

$$J(w) = \frac{1}{2}\sum_{i=1}^{n}[y^{(i)} - \Phi(z^{(i)})]^2$$



A geometric interpretation of L2 regularization

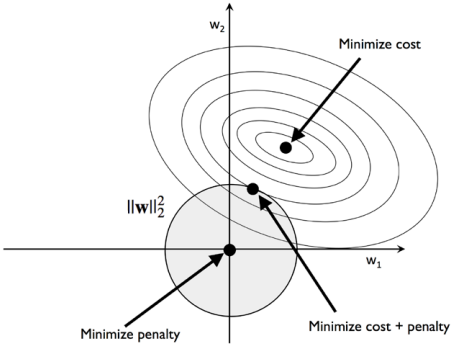
Khi chúng ta thêm thành phần L2 regularization vào cost function, ta có:

$$J_2(w) = J(w) + \lambda \sum_{j=1}^{m} w_j^2 = \frac{1}{2} \sum_{i=1}^{n} [y^{(i)} - \Phi(z^{(i)})]^2 + \lambda \sum_{j=1}^{m} w_j^2$$

Bài toán 1: tìm w để $J_2(w)$ min.
Bài toán 2: tìm w để $J(w)$ min, với điều kiện $\sum_{j=1}^{m} w_j^2 = t$.
2 bài toán trên là tương đương với nhau.

A geometric interpretation of L2 regularization

Ta có đồ thị của bài toán trên:

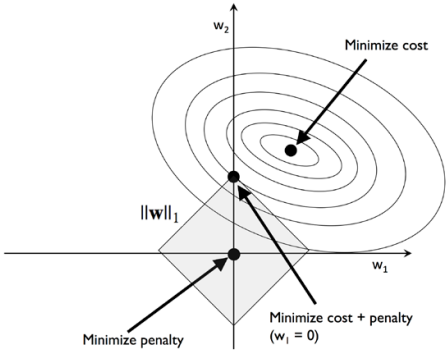


Mục lục

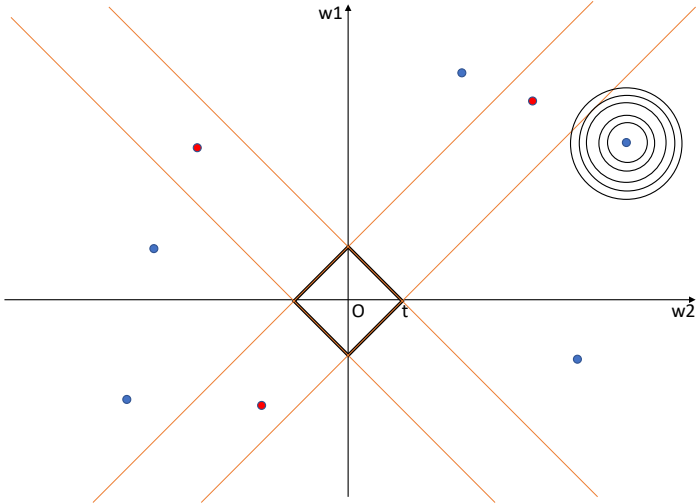
Sparse solutions with L1 regularization

$J(w) = \frac{1}{2} \sum_{i=1}^{n} [y^{(i)} - \Phi(z^{(i)})]^2$

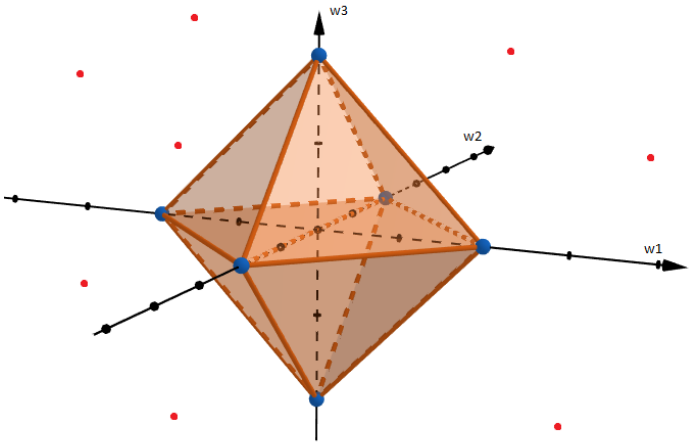
Bài toán: tìm w để $J(w)$ min, với điều kiện $\sum_{j=1}^{m} |w_j| = t$



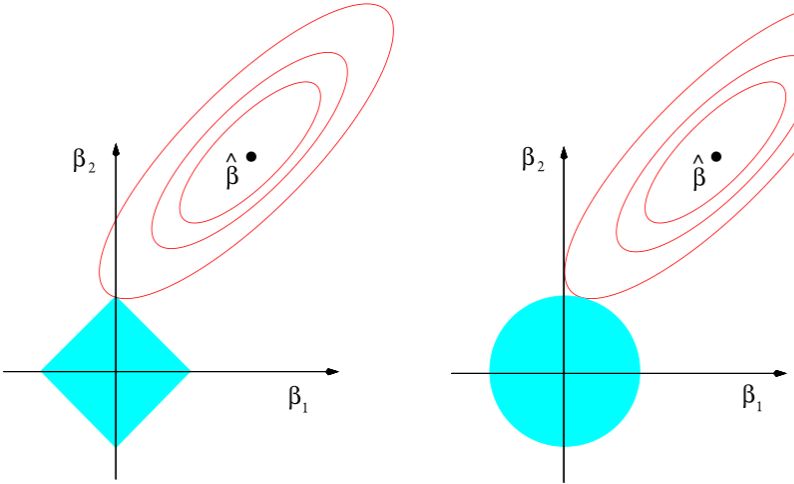
Sparse solutions with L1 regularization



Sparse solutions with L1 regularization

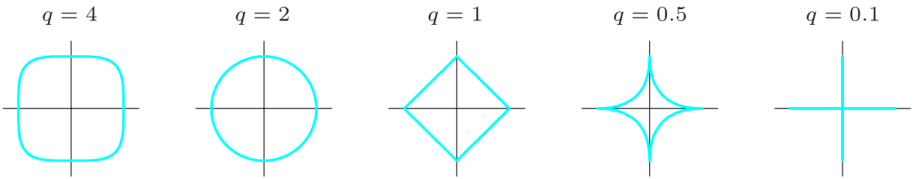


Sparse solutions with L1 regularization



Sparse solutions with L1 regularization

Tổng quát về regularization: $\sum_{j=1}^{m} \left| w_j \right|^q$



Sparse solutions with L1 regularization

Mục lục

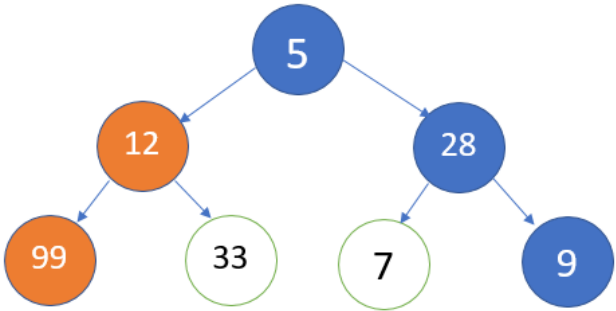
Sequential feature selection algorithms

Có hai kỹ thuật **dimensionlity reduction** là: **feature selection** và **feature extraction**.
- **feature selection**: chọn một tập con của tập các feature ban đầu (nếu có m features, thì ta có tổng cộng 2^m tập con).
VD: tập feature ban đầu là $\{x_1, x_2, x_3\}$, ta chọn tập con $\{x_1, x_3\}$
- **feature extraction**: dựa vào tập feature ban đầu, ta tạo nên một tập feature mới.
VD: tập feature ban đầu là $\{x_1, x_2, x_3\}$, ta tạo một tập feature mới là $\{y_1, y_2\}$, trong đó $y_1 = x_1 + x_2, y_2 = x_3$.

Bài này sẽ trình bày những giải thuật **feature selection**.

Sequential feature selection algorithms

Những giải thuật **sequential feature seclection** là một họ của những giải thuật **greedy search** (hay còn gọi là giải thuật tìm kiếm tham lam).



Sequential feature selection algorithms

Một giải thuật sequential feature selection kinh điển đó là **sequential backward selection** (SBS) (ngoài ra còn có SFS, LRS, BDS, SFFS).
Các bước tiến hành SBS:

1. khởi tạo số lượng feature mong muốn giữ lại là k_d, khởi tạo $k = d$, trong đó d là số lượng feature của tập feature ban đầu X_k.
2. loại bỏ feature x_i trong tập X_k dựa vào một objective function $J(X_k, x_i)$, $k := k - 1$.
3. quay lại bước 2 nếu k $>$ k_d.

Vậy objective function $J(X_k, x_i)$ có công thức như thế nào?

Sequential feature selection algorithms

Có 2 nhóm objective function cho bài toán feature selection, đó là: **Filters** và **Wrappers**
- Filters: đánh giá tập con bằng thông tin chứa trong nó.
- Wrappers: đánh giá tập con bằng độ chính xác của mô hình trên tập con đó.

Bài trình bày sẽ mô tả chi tiết các bước thực hiện giải thuật SBS với wrapper objective function.

Sequential feature selection algorithms

Khởi tạo số lượng feature mong muốn giữ lại là k_d, khởi tạo $k = d$

1. Tách data làm 2 tập: tập train và tập test. Tập train là X_k (trong đó k là số lượng feature).
2. Lần lượt tính độ chính xác của các tập con $X_k \setminus \{x_i\}$ (nghĩa là số feature của tập con ít hơn 1 feature so với tập cha), sau đó giữ lại tập con có độ chính xác cao nhất, gán $k := k - 1$.
3. Nếu $k > k_d$ thì quay lại bước 2.

Lưu ý bước 2: Ta tách tập con thành 2 tập là tập train và tập validation. Trong đó tập train dùng để train model, tập validation để tính độ chính xác của mô hình vừa train.