Ngày 2

Huỳnh Vũ Khôi Nguyên

1 Linear Regression

Ví dụ về việc dự đoán giá nhà: Ta có một tập các bức ảnh về căn nhà. Mục tiêu của ta là dự đoán giá nhà dựa trên các bức ảnh đó.
Theo mô hình chung ta đã nói ở bài trước, mục tiêu của ta là tìm hàm $f : \mathbb{R}^{m\times n\times 3} \to \mathbb{R}$ ($\mathbb{R}^{m\times n\times 3}$ là không gian các bức ảnh màu và $\mathbb{R}$ là không gian giá nhà).

Ở đây ta giả sử đã tìm được các basis function và từ đó có được một coordinate vector z hay là việc g_1 trong mô hình chung đã được trích xuất. Việc của chúng ta là từ coordinate này đưa ra giá của căn nhà. Một cách đơn giản để làm việc này là tìm một ánh xạ tuyến tính $\mathbb{R}^d \to \mathbb{R}$. Đây là ánh xạ g_2 trong mô hình chung của chúng ta
Các bài toán có đầu ra là một số như thế này được gọi là bài toán regression.
Ta viết lại bài toán dưới mô hình toán học như sau:
Cho trước một dataset với N điểm dữ liệu $D = \left\{z^i, y^i\right\}_{i=1}^N$, cần tìm ánh xạ tuyến tính f phù hợp nhất với mô hình của chúng ta.

$$\begin{aligned} f : \mathbb{R}^n &\to \mathbb{R} \\ z^\intercal &\to \widehat{y} \end{aligned}$$

Trong đó $z^\intercal$ là coordinate vector và y là số thực biểu thị giá nhà.
Gọi w là ma trận biểu thị ánh xạ tuyến tính f thì mối quan hệ có thể được viết lại thành $f(z) = z^\intercal w = \widehat{y}$.
Ta cần tìm f tối ưu để giá trị đầu ra gần với giá trị quan sát được nhất tương đương với việc tìm w^* sao cho

$$w^* = \underset{w}{argmin} \sum_{i=1}^{N} d(\widehat{y}, y)$$

với d là một chuẩn đánh giá.

2 Classification

Ta xét ví dụ về bài toán phân loại ảnh với đầu vào là các bức ảnh và đưa ra bức ảnh thuộc loại nào trong d nhóm ta đã chon trước.

Như trên, ta cũng giả sử rằng đã tìm được các basis function và có được coordinate vector z hay hàm g_1 trong mô hình chung đã được xác định.

Definition. Vector xác suất (probability vector) $p = [p_1, p_2, ..., p_d]^\intercal$ được gọi là một vector xác suất nếu hai điều kiện sau đây thỏa mãn:

- $p_i \in [0, 1]\ \forall i$
- $\sum_{i=1}^{d} p_i = 1$

Do bài toán yêu cầu phân loại ảnh nên ta sẽ tìm một ánh xạ chuyển coordinate vector như trên về một vector xác suất (probability vector) biểu thị sự tin tưởng của ta rằng ảnh có bao nhiêu phần thuộc loại nào. Đây là ánh xạ g_2 ở mô hình tổng quát.
Sau bước này ta thu được một probability vector $\widehat{y} = [\widehat{y}_1, \widehat{y}_2, ..., \widehat{y}_d]^\intercal$ với mỗi $\widehat{y}_i$ là mức độ tin tưởng của ta rằng bức ảnh có bao nhiêu phần thuộc vào nhóm thứ i.
Để có thể so sánh với probability vector $\widehat{y}$ ở trên, ta cũng phải chuyển output của ta về một probability vector. Ta thực hiện việc này bằng one-hot encoding. Nếu xếp bức ảnh vào nhóm thứ i thì one-hot encoding sẽ là một vector $y \in \mathbb{R}^d$ có phần tử 1 duy nhất nằm ở vị trí thứ i. Đây là bước embedding trong mô hình tổng quát.
Đến đây, ta so sánh y và $\widehat{y}$ qua $d(\widehat{y}, y)$ với d là một chuẩn đánh giá.

Một chuẩn đánh giá hiệu quả được sử dụng trong các bài toán classification là cross-entropy. Cross-entropy giữa y và $\widehat{y}$ được tính bằng:

$$-\sum_{i=1}^{d} p_{y_i} log\,(p_{\widehat{y}_i})$$

Ở bước chuyển coordinate vector về probability vector ở trên (ánh xạ g_2), ta có thể dùng softmax function.
Giả sử $z = [z_1, z_2, ..., z_d]^\intercal$ là coordinate vector. Áp dụng softmax function lên mỗi z_i ta được $\widehat{y}_i$:

$$\widehat{y}_i = softmax_i = \frac{\sigma(z_i)}{\sum_{j=1}^{d} \sigma(z_j)}$$

Khi đó từ biểu thức nhận thấy rằng $\widehat{y}_i \in [0, 1],\ \forall i$ và $\sum_{i=1}^{d} \widehat{y}_i = 1$ nên $\widehat{y} = [\widehat{y}_1, \widehat{y}_2, ..., \widehat{y}_d]^\intercal$ là probability vector.

Ta coi việc phân loại một đối tượng vào nhóm thứ i là việc xuất ra output là probability vector của các class (h_1 trong framework). Ta giả sử đã có được coordinate vector z. Từ coordinate vector này, ta cần chuyển về probability vector để có thể so sánh với probability vector của các class (g_2 trong framework). Đánh giá này có thể thực hiện một cách hiệu quả bằng cross-entropy.

Đối với các mô hình linearly seperable thì $g_2 = \gamma(Wz)$ với W là phép biến đổi tuyến tính và γ là một hàm phi tuyến. Logistic regression và Softmax regression là các mô hình như vậy. Đối với Logistic regression thì γ chính là hàm sigmoid σ và số lượng class là 2. Đối với Softmax regression thì γ chính là hàm softmax như đã nói ở trên.
Đối với các mô hình linearly non-seperable thì $g_2 = \gamma_1(\gamma_2(...\gamma_n(z)))$ với γ_i là hàm phi tuyến