

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



BÀI TẬP

MÔN CÁC KỸ THUẬT HỌC SÂU

VÀ ỨNG DỤNG

KHOA: KHOA HỌC MÁY TÍNH

HOMEWORK: GIẢI THÍCH HÀM LOSS CỦA LOGISTIC REGRESSION

Giảng viên: Nguyễn Duy Khánh

Nhóm thực hiện:

1. **Trương Văn Khải– 21520274**

Đề: Các em hãy giải thích tại sao sử dụng hàm Binary Cross Entropy mà không sử dụng MSE cho hàm loss của mô hình phân lớp nhị phân với Logistic Regression.

Bài làm:

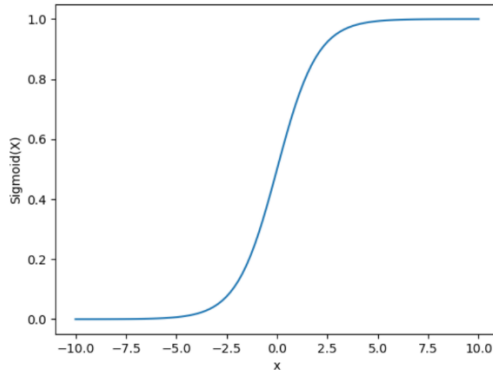
Lý do 1: Chúng ta giải quyết các vấn đề phân lớp này từ một cái nhìn xác suất. ($y_{pred} \in (0; 1)$).

- Vì Logistic Regression là 1 bài toán classification nên giá trị trả về sẽ nằm trong khoảng từ $[0;1]$.
- Nếu ta dùng MSE là hàm loss cho bài toán Logistic Regression, ta sẽ gặp phải vấn đề sau:
 - o Công thức MSE:

$$\operatorname{argmin}(\theta) \frac{1}{2n} \sum_{i=1}^n (f(x_i) - y_i)^2$$

- o Vì vậy, giá trị của $f(x) = \theta^T x$ sẽ phụ thuộc vào x nên giá có giá trị từ $[-\infty; +\infty]$. Khi đó hàm Loss sẽ phụ thuộc nhiều vào $f(x_i)$ (ví dụ: $f(x_i)$ giả sử có giá trị là 1000, còn y sẽ có giá trị là 1) thì khi đó mô hình sẽ tập trung vào vấn đề tìm θ để giảm giá trị của $f(x_i)$ xuống thay vì phải làm công việc chính của nó là xấp xỉ 2 giá trị $f(x_i)$ và y_i .
- o Giải pháp đặt ra là phải tìm được 1 hàm để ép giá trị của $f(x_i)$ về nhỏ hơn, lúc đó ta cần sử dụng hàm $\operatorname{sigmod}(x) = \frac{1}{1+e^{-x}}$ với đồ thị bên dưới để có thể đưa giá trị $f(x_i)$ giới hạn về khoảng $[0;1]$.

Figure 1



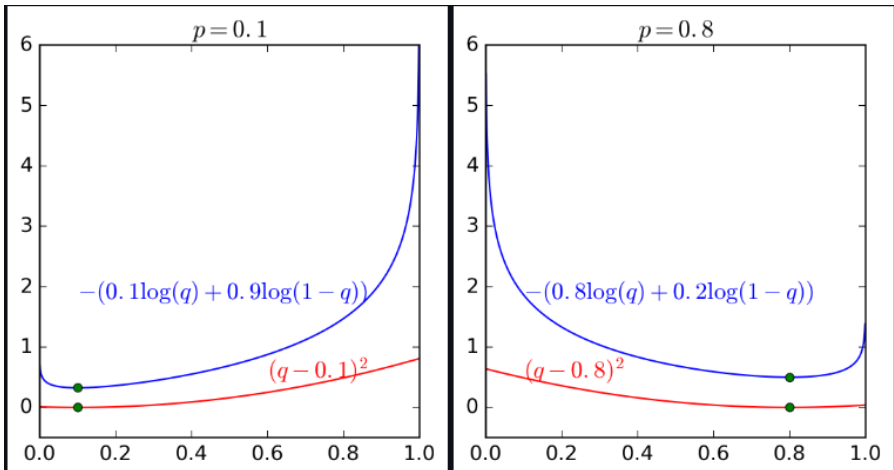
Lý do 2: Cross Entropy giữa 2 phân phối p và q được định nghĩa là:

$$H(p, q) = E_p[-\log q]$$

Còn nếu với p và q là rời rạc, công thức trên sẽ được viết dưới dạng:

$$H(p, q) = -\sum_{i=1}^C p_i \log q_i$$

Để hiểu rõ hơn ưu điểm của hàm Cross Entropy và hàm bình phương khoảng cách thông thường, chúng ta sẽ xem hình dưới đây. Đây là ví dụ trong trường hợp $p = 0.8$ và $p = 0.1$.



Hình trên thể hiện sự khác nhau giữa 2 hàm Cross Entropy và hàm bình phương khoảng cách. Các điểm màu xanh lục thể hiện các giá trị nhỏ nhất của mỗi hàm.

Nhận xét:

- Giá trị của cả hai hàm số đạt được khi $p = q$ thể hiện ở các điểm màu xanh lục
- Quan trọng hơn hết, hàm Cross Entropy mang giá trị rất cao (tức là loss rất cao) khi p và q ở xa nhau. Về mặt tối ưu, hàm Cross Entropy sẽ cho nghiệm gần p hơn vì những nghiệm ở xa sẽ bị trừng phạt rất nặng. \Rightarrow Tính chất này khiến cho hàm Cross Entropy được sử dụng rất rộng rãi khi tính khoảng cách giữa hai phân phối xác suất.
- Trong khi đó, sự chênh lệch giữa các loss dù cho ở gần hay ở xa của $(p - q)^2$ là không đáng kể.

Chú ý: Hàm Cross Entropy không có tính đối xứng, tức $H(p, q) \neq H(q, p)$. Chính vì vậy, khi sử dụng Cross Entropy trong bài toán Logistic Regression, p thường là đầu ra thực sự vì đầu ra thực sử chỉ có 1 thành phần bằng 1, còn lại bằng 0, q thường là đầu ra dự đoán khi mà không có xác suất tuyệt đối bằng 1 hoặc tuyệt đối bằng 0 cả.

Lý do 3: Áp dụng Binary Cross Entropy trên bài toán phân lớp sẽ hội tụ nhanh hơn. ($\mathcal{L}_{MSE} \in [0; 1]$, $\mathcal{L}_{BCE} \in [0; +\infty)$).