Softmax Regression

Mô hình hồi quy Logistic có thể được tổng quát cho bài toán đa lớp mà không cần huấn luyện và tổ hợp các bộ phân lớp nhị phân. Cách thực hiện này gọi là Hồi quy Softmax hay Hồi quy Logistic đa thức (Multinomial Logistic Regression).

Ý tưởng: Với dữ liệu mới x, hồi quy Softmax ước lượng xác suất để x thuộc lớp k như sau:

B1: Tính các điểm $S_k(x)$ cho lớp thứ k

Equation 4-19. Softmax score for class k

$$s_k(\mathbf{x}) = \mathbf{x}^\mathsf{T}\boldsymbol{\theta}^{(k)}$$

Tất cả vector tham số Theta(k) được lưu thành hàng của ma trận tham số A.

B2: Ước lượng xác suất cho mỗi lớp bằng cách sử dụng hàm softmax (Softmax function-còn gọi là hàm mũ chuẩn hóa- Normalized exponential)

Equation 4-20. Softmax function

$$\hat{p}_k = \sigma(\mathbf{s}(\mathbf{x}))_k = \frac{\exp\left(s_k(\mathbf{x})\right)}{\sum_{j=1}^{K} \exp\left(s_j(\mathbf{x})\right)}$$

K= số lớp

S(x)= vector chứa các điểm số cho mỗi lớp

Sigma$(s(x))_k$ = xác suất x thuộc lớp k

Tương tự bộ phân lớp hồi quy logistic, bộ phân lớp hồi quy Softmax dự đoán x thuộc về lớp có xác suất lớn nhất

Equation 4-21. Softmax Regression classifier prediction

$$\hat{y} = \underset{k}{\operatorname{argmax}}\ \sigma(\mathbf{s}(\mathbf{x}))_k = \underset{k}{\operatorname{argmax}}\ s_k(\mathbf{x}) = \underset{k}{\operatorname{argmax}}\ \left(\left(\boldsymbol{\theta}^{(k)}\right)^\mathsf{T}\mathbf{x}\right)$$

Để ý: Mô hình hồi quy Softmax chỉ dự báo 1 lớp tại một thời điểm, tức là thuộc lớp bài toán phân đa lớp, không phải bài toán đa nhãn (multioutput)

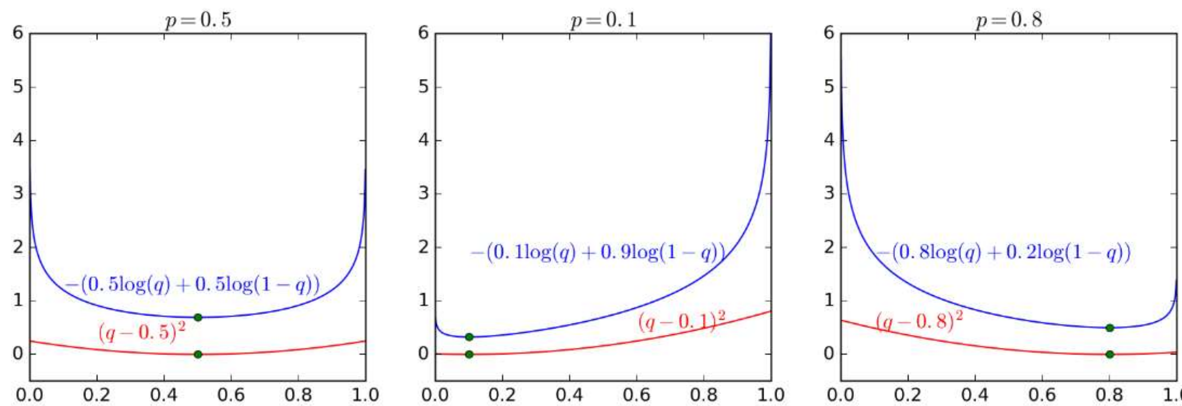
VD:

Nhân diện loại cây: Bài toán đa lớp

Nhận diện nhiều người trong một bức hình: Bài toán đa nhãn

Hàm cross-entropy

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



Hình 4: So sánh giữa 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 nhỏ nhất của cả hai hàm số đạt được khi $q=p$ tại hoành độ của các điểm màu xanh lục.
- Hàm cross entropy nhận giá trị rất cao (tức loss rất cao) khi q ở xa p. Trong khi đó, sự chênh lệch giữa các loss ở gần hay xa nghiệm của hàm bình phương khoảng cách $(q-p)^2$ là không đáng kể. Về mặt tối ưu, hàm cross entropy sẽ cho nghiệm *gần* với p hơn vì những nghiệm ở xa bị *trừng phạt* rất nặng.

Hai tính chất trên đây khiến cho cross entropy được sử dụng rộng rãi khi tính khoảng cách giữa hai phân phối xác suất.

Hàm mất mát của Softmax regression

Equation 4-22. Cross entropy cost function

$$J(\boldsymbol{\Theta}) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)}\log\left(\hat{p}_k^{(i)}\right)$$

Với $y_k^{(i)}$ (nhận giá trị 0 hoặc 1) là xác suất quan sát thứ i thuộc lớp k.

Equation 4-23. Cross entropy gradient vector for class k

$$\nabla_{\boldsymbol{\theta}^{(k)}} J(\boldsymbol{\Theta}) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{p}_k^{(i)} - y_k^{(i)}\right)\mathbf{x}^{(i)}$$

```
X = iris["data"][:, (2, 3)]  # petal length, petal width
y = iris["target"]

softmax_reg = LogisticRegression(multi_class="multinomial",solver="lbfgs", C=10)
softmax_reg.fit(X, y)
```

Figure 4-25 shows the resulting decision boundaries, represented by the background colors. Notice that the decision boundaries between any two classes are linear. The figure also shows the probabilities for the *Iris versicolor* class, represented by the curved lines (e.g., the line labeled with 0.450 represents the 45% probability boundary). Notice that the model can predict a class that has an estimated probability below 50%. For example, at the point where all decision boundaries meet, all classes have an equal estimated probability of 33%.

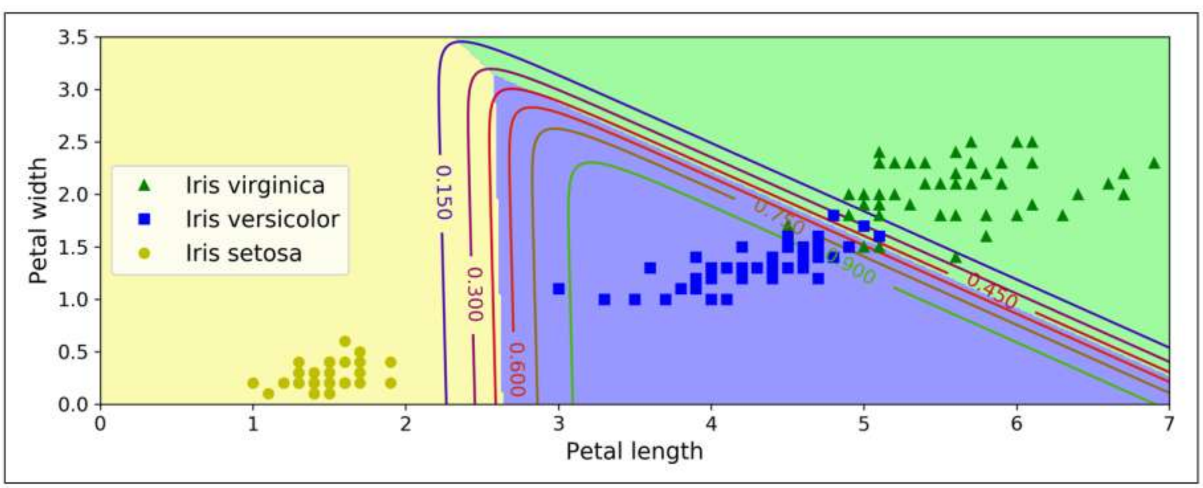


Figure 4-25. Softmax Regression decision boundaries