

Classification Problem

Guorui Shen

guorui233@outlook.com

May 2, 2020

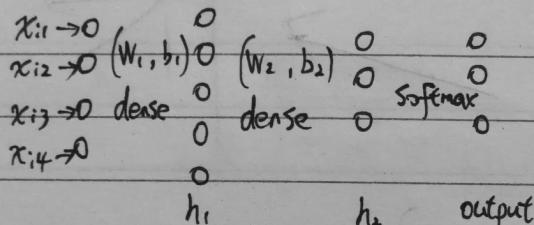
① Take a point in Iris dataset as example:

$$x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$$

$$y_i = (1, 0, 0) \text{ or } (0, 1, 0) \text{ or } (0, 0, 1)$$

Let's say $y_i = (1, 0, 0)$,

② Using the neural network below to predict it:



$$y_i = \sigma(x_i W_1 + b_1) W_2 + b_2$$

$$h_2 = \sigma(x_i W_1 + b_1) W_2 + b_2$$

$$\text{output} = \text{softmax}(h_2)$$

$$= \left[\frac{h_{21}, h_{22}, h_{23}}{\sum_{i=1}^3 h_{2i}} \right] \in \mathbb{R}^{1 \times 3}$$

Question 1. Why don't we simply use h_2 to approximate y_i ?Answer 1. We can directly use h_2 to approximate y_i . Then use $\|h_2 - y_i\|_2^2$

to calculate the loss. But if we use cross entropy, defined by

$$\text{cross_entropy}(p, q) = - \sum_{i=1}^n p_i \log q_i, \quad p_i \in [0, 1], \quad q_i \in (0, 1)$$

as the loss function, then we can't use h_2 to estimate y_i .Since the entry of h_2 could be negative.

Question 2. What's the relationship between cross entropy & prediction error?

Answer 2. The smaller the cross entropy, the better the model parameter and therefore the smaller the prediction error.

Conclusion: If the cross entropy loss is used, then we have to use softmax at the final layer of the neural network.

③ The drawback of using label $\{0, 1, 2, \dots, C-1\}$, C classes in total.

In such a non-one-hot labelling way, class i & class $i+1$ are then thought to have a similarity between them. This doesn't make sense for datasets such as Iris dataset

Question 3. When to use label $\{0, 1, 2, \dots, C-1\}$ & when to use one-hot label $\{[1, 0, \dots, 0], [0, 1, 0, \dots, 0], \dots, [0, \dots, 0, 1]\}$?

Answer 3. If two adjacent classes have some relationship, such as

$$\begin{aligned} &\{ \text{no risk}, \text{low risk}, \text{middle risk}, \text{high risk} \} \\ &= \{ 0, 1, 2, 3 \} \end{aligned}$$

then use non-one-hot.

~~Else~~ Else use one-hot.

④ Classic classification problems / Methods.

problem :	Binary classification problem. Iris dataset.	• Minst handwriting test.
-----------	---	---------------------------

- | | |
|-----------|--|
| Methods : | <ul style="list-style-type: none"> SVM $\{-1, 1\}$ Logistic regression $\{0, 1\}$ Multiclass classification |
|-----------|--|