

Multiclass Classification Problem

we use a one-vs-rest approach

- a binary classification problem is fit for each class, treating that class as a **positive class** and all other classes as a **negative class**
- during the prediction, the class with the highest predicted probability among all binary classifiers is chosen as the final prediction

Example - 3 classes

- $c1(x) = 0.7$
 - $c2(x) = 0.3$
 - $c3(x) = 0.8$
- here the third class is selected as it has the largest probability

Multinomial (multiclass) loss

- logistic regression model is optimized for multinomial loss
- the entire probability distribution across all classes is considered jointly during approximation
- the softmax function is used to calculate probabilities across all classes and the model is trained to minimize the multinomial loss

"Pasted image 20240205133445.png" could not be found.

given some user-supplied weights, we can find the probability of each class.

"Pasted image 20240205134032.png" could not be found.

- we label how much each one-hot vector (x) affects each class (weighting)
- then sum over each vector in each class
- pass through soft-max
- select class with highest resultant probability

Cross-Entropy Loss as cost function

how do you measure the "distance" between 2 distributions?

- information theoretic measure of distance between two distributions

- if 2 distributions are the same, then 0 while is has a nonnegative value