Cross Entropy

Table of Contents

Entropy 2
Prediction Model 2

Entropy

In the previous videos we have seen that the performance of a strategy is measured by using a loss function. In this document, you will understand how to calculate this loss. We will use the cross-entropy function to do the same.

Prediction problem: Let us say that we want to predict the trend of the returns for the next day. We want to predict if the next day's return would be within the first standard deviation or the second standard deviation of returns. So, our prediction problem can be broken down into four classes:

$\boldsymbol{\sigma}$: Returns are between 0 and 1st standard deviation

-$\boldsymbol{\sigma}$: Returns are between 0 and -1st standard deviation

2$\boldsymbol{\sigma}$: Returns are between 1st and 2nd standard deviation

-2$\boldsymbol{\sigma}$: Returns are between 1st and 2nd standard deviation

Prediction Model: Let us say that we have a prediction model for which the input is RSI (Relative Strength Index. In the table below, the model calculates the probability of 'returns' belonging to the class of '2$\boldsymbol{\sigma}$'. The column 'output' is given in terms of probabilities as seen below. If the predicted probability is greater than 0.50 for a class, then the data point is assigned to that class. For example, in the table above, data point 1 is predicted to be in 2$\boldsymbol{\sigma}$ class, data point is predicted to be not in 2$\boldsymbol{\sigma}$ class.

Data	Input	Output probability	Actual class	Predicted Class	Was the prediction Correct (1 if correct or 0 if wrong)
1	70	0.65	2$\boldsymbol{\sigma}$	2$\boldsymbol{\sigma}$	1
2	50	0.25	2$\boldsymbol{\sigma}$	2$\boldsymbol{\sigma}$	1
3	30	0.10	-$\boldsymbol{\sigma}$	2$\boldsymbol{\sigma}$	0

The table above can be understood as follows:

A. For the first data point, where the RSI was 70, the classifier predicted that the next day's return would belong to the class '2$\boldsymbol{\sigma}$' and it turned out to be correct, hence we have a '1' in the last column. The classifier predicted this with a probability of 65%.
B. For the second data point, where the RSI was 50, the classifier predicted the same with a probability of 25%.
C. For the last data point, where the RSI was 30, the classifier predicted same with a probability of 10%.

Now to measure the loss of the classifier we use the cross-entropy function given as follows:

$$\sum_i p_i \times log_2 \frac{1}{\dot{p}_i}$$

Where,

P_i is the actual probability of the data point belonging to a class i

$\dot{P}_i$ is the predicted probability of the data point belonging to a class i

For the first datapoint, the actual probability of it belonging to 2$\boldsymbol{\sigma}$ is 100%, and the predicted probability for this 65%. Replacing these two in the entropy equation we get:

$$100\% * \log_2(1/0.65)$$

Similarly, we can calculate the total loss of the classifier over three data points as follows:

$$\text{Loss} = 1 * \log_2(1/0.65) + (1) * \log_2(1/0.25) + (1-0)*\log_2(1/(1-0.1))$$
$$= -1.922$$

Had the prediction for data point 3 been correct, with a probability of 0.75, then this loss value would be -2.1, which is lesser than what is achieved now. . From this we can infer that as the predictions become more accurate, along with higher probability, the loss will keep reducing.

The total s loss calculation uses all the data points. As the number of data points increases, the sum of the losses would become more negative, but does not mean that the model is performing better. To counter this, we measure the total loss as the average loss per example. In the example above the average loss would be -1.922/3 =-0.64. This is the final cross entropy loss value for the classifier from the above example.