

Concordant & Discordant Pairs for Logistic Regression

When we think of model evaluations metric - we think of accuracy , recall , precision or f1_score. Today lets see some other well -known methods.

Welcome to Concordant & Discordant Pairs

Lets take a sample data set which has 1000 sample of which we have 100 events (Target =1) and 900 Non events (Target =0) .

Now when we do predictions we get probabilities for each sample for e.g. [0.3,0.7]. i.e. the probabilities of this sample being Non Event is 0.3 and Event is 0.7.

Step 1 : Pair Creation

Since we have 900 Non Events & 100 Events - we will have $900 \times 100 = 90000$ pairs, where each pairs comprises of 2 values (X, Y) where X represents probability of Event and Y represents probability of Non Event.

Step 2 : Concordant Pairs

All the pairs where $X > Y$ are termed as Concordant pairs. for e.g. we have 80000 concordant pairs.

Step 3 : Discordant Pairs

All the pairs where $X < Y$ are termed as Discordant pairs for e.g. we have 7000 discordant pairs.

Step 4: Tier Pairs

All the pairs where $X = Y$ are termed as Tier Pairs for e.g. we have 3000 tied pairs.

Count the total number of Concordant pairs and divided by the total number of the pairs i.e.

Concordance ratio = $80000/90000 = 0.8888$ or 88.88%

Higher the concordance ratio - better is the model.

****Somers D: This is (% concordant pair - % discordant pair). ****

Somers D = $80000-70000/90000 = 81.11\%$

Higher Somers D indicates a better model