

theory7

by Kondrashov Anton

March 16, 2017

1 Task

Task description says that while in crisis bank tends to decline loans more often. That means more negatives and less positives. Given that bank is choosing more reliable clients we can suppose that ratio of true positives to false positives at least stays the same. Or becomes larger. Then the bank has lower FPR value and that means that it should choose the curve with high TPR while having low FPR. That one is blue.

2 Task

As the task is binary classification, there are only two possible classes of object. So, if we have error of 90% then we may just invert our model's answers and that will result in 10% error. That trick doesn't work if we have 50% error.

3 Task

The LOO is computationally expensive and has high variance. But it gives more complete estimation of error. So

1. On small dataset it's better to use LOO as it won't require much time because of the size of dataset, but will give better estimation.
2. On very large dataset it's better to use k-fold with small k as it won't take much time for computations and variance is lower.