Report Lazyfca

Samarkhanov Ratmir

My solution has 3 different implementations for task:

A training data is divided into two parts - with a positive target feature and a negative one.

Metrics:

- o True Positive
- o True Negative
- o False Positive
- o False Negative
- o True Positive Rate
- o True Negative Rate
- o Negative Predictive Value
- o False Positive Rate
- o False Discovery Rate
- o Accuracy
- o Precision
- o Recall

TP - correct positive if we predicted a positive class and the real value of the class is positive (and the ratio of positive to negative exceeds the threshold)

FP - wrong positive if we predicted a positive class, and the real value of the class is negative (and the ratio of positive to negative exceeds the threshold)

TN - correct negative if we predicted a negative class and the real value of the class is negative (and the ratio of negative to positive exceeds the threshold)

FN - incorrect negative, if we predicted a negative class, and the real value of the class is positive (and the ratio of negative to positive exceeds the threshold)

First:

We use norm sum of intersection of features:

$$positive = \frac{sum \ |g' \cap g_i+|}{|G_+|}$$

$$negative = \frac{sum \ |g' \cap g_i\neg|}{|G_-|}$$

Where $|g' \cap g_i^+|$ – мощность пересечения

Our answer is where sum is bigger.

With this algorithm I get average accuracy ~ = 0.652

Second:

We just check count of matches from negative or positive clas.

And then get pos/neg where count is bigger.

With this algorithm: I get accuracy ~=0.6388

Metric (AVG)	Algorithm 2
TP	350
TN	200
FP	97
FN	213
TPR	0.6211
TNR	0.6721
FPR	0.1738
NPV	0.4849
FDR	0.2182
ACCURANCY	0.63881
PRECUSION	0.7817
RECALL	0.6211

Third:

It's much slower than two first but much more precisely.

In this algorithm we use similar idea from first but now we use parameter 'votes' with possibility of solve will we add vote or not. If number of votes bigger than threshold – we'll add.

Which number of votes of class positive/negative is bigger – win.

While testing best number of votes became 8.

With this algorithm with 8 threshold I get average accuracy ~= 0,97977