Train In Data

MetaCost

MetaCost

MetaCost is a procedure to make a cost insensitive algorithm, cost sensitive

- It can be applied to any algorithm
  - Whether it returns probabilities or classes.

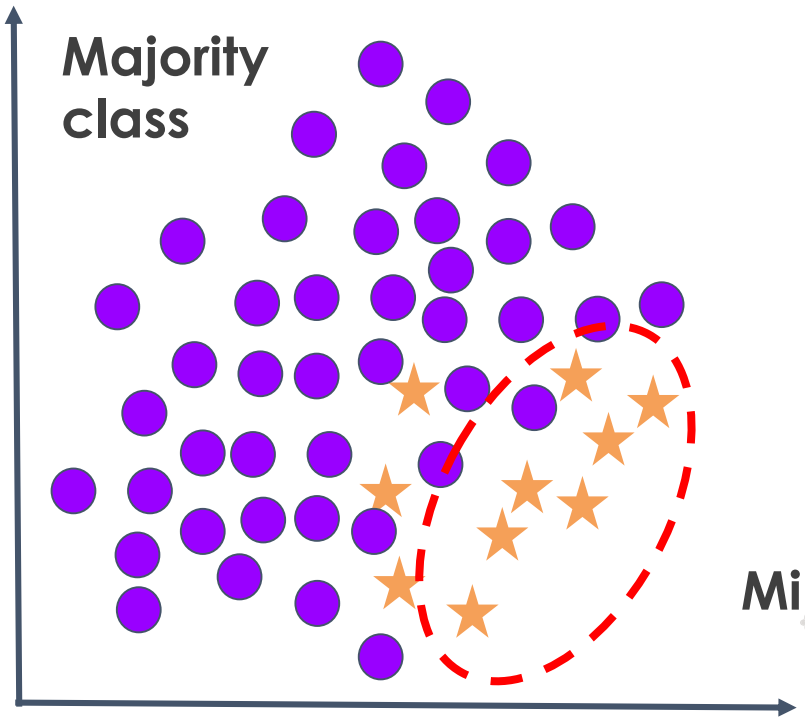
Conditional Risk

The Bayes optimal prediction for an observation is the class that minimises the conditional risk:

$$R(i|x) = \sum_{j=1}^{M} P(j|x) \cdot C(i, j),$$

Prediction boundaries

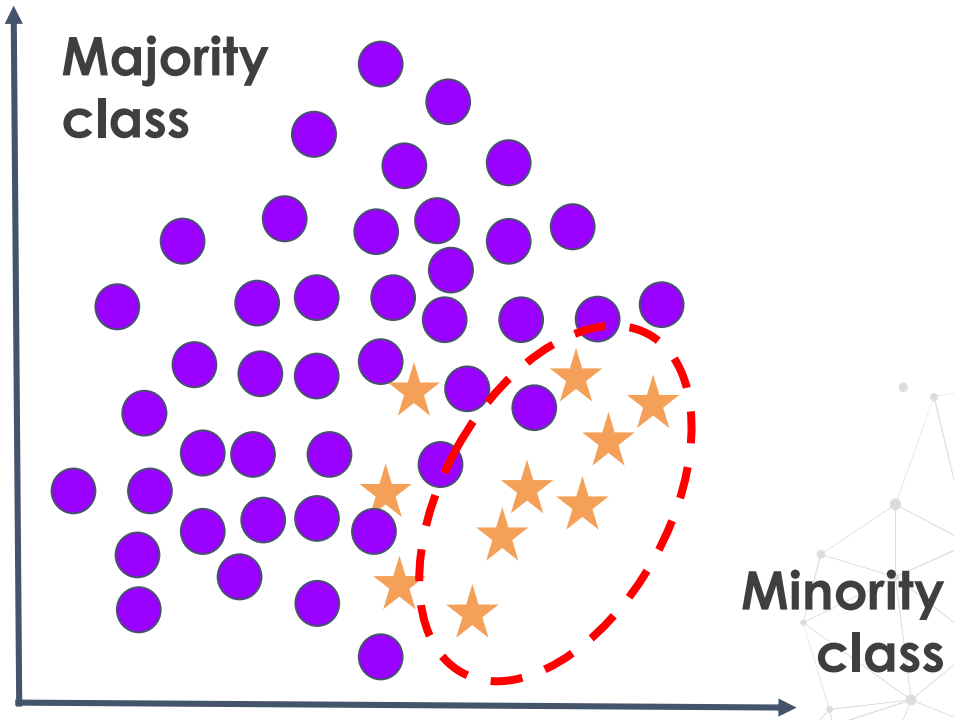
The role of a classifier is to find the boundaries that separate the classes.



Prediction boundaries – cost sensitive

Cost of misclassification of the minority class is higher.

$$R(i|x) = \sum_{j=1}^{M} P(j|x) \cdot C(i, j),$$

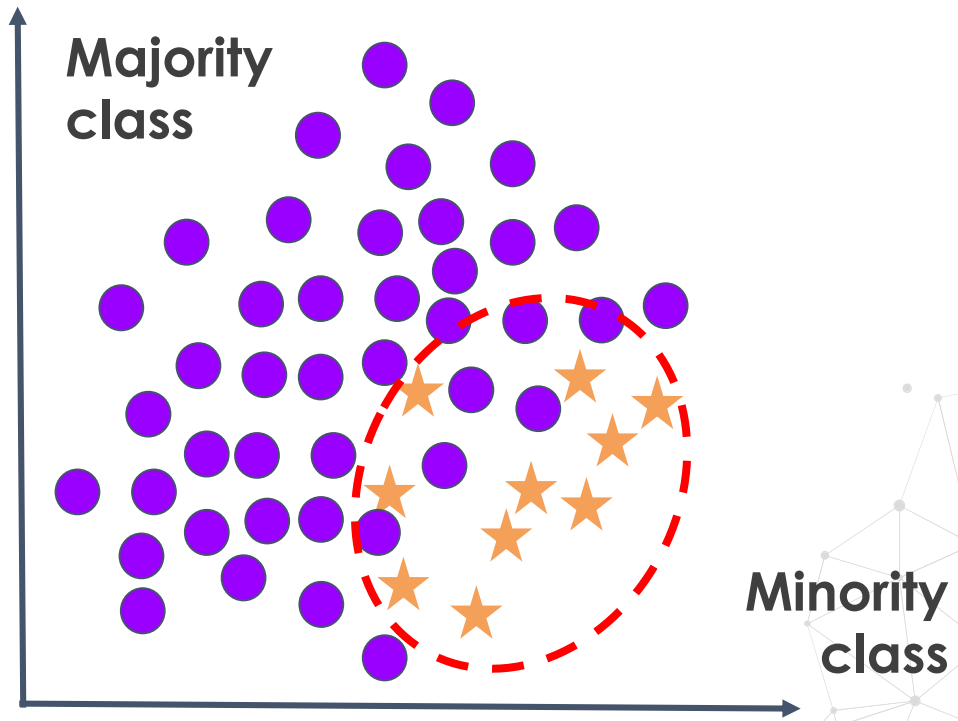


Prediction boundaries – cost sensitive

Cost of misclassification of the minority class is higher.

Thus, the region where we should predict 1, or the minority class expands

$$R(i|x) = \sum_{j=1}^{M} P(j|x) \cdot C(i, j),$$

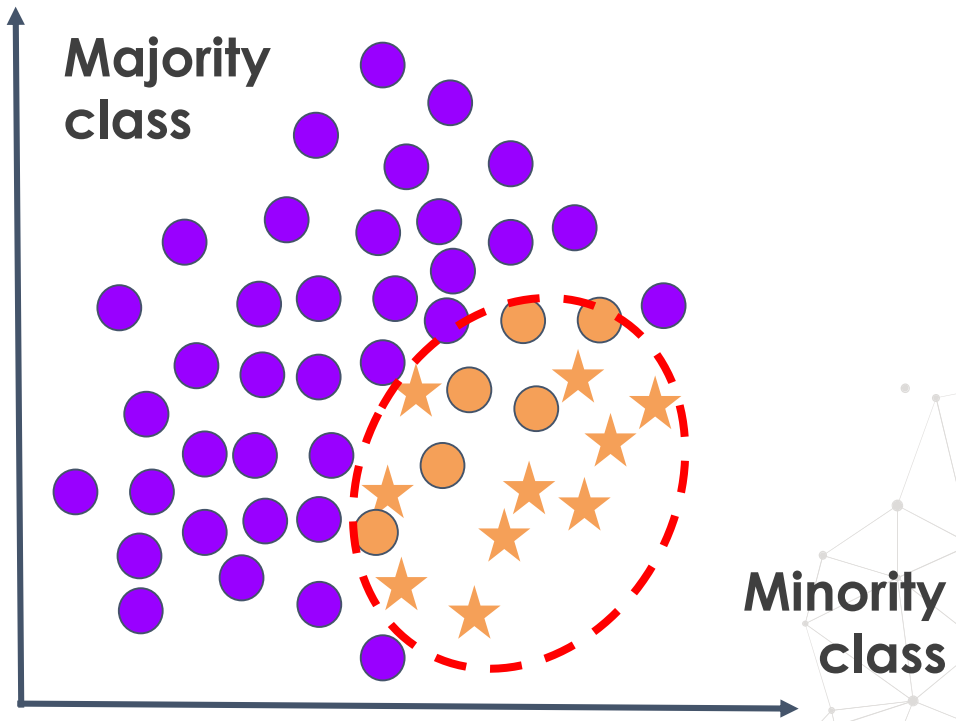


Prediction boundaries – cost sensitive

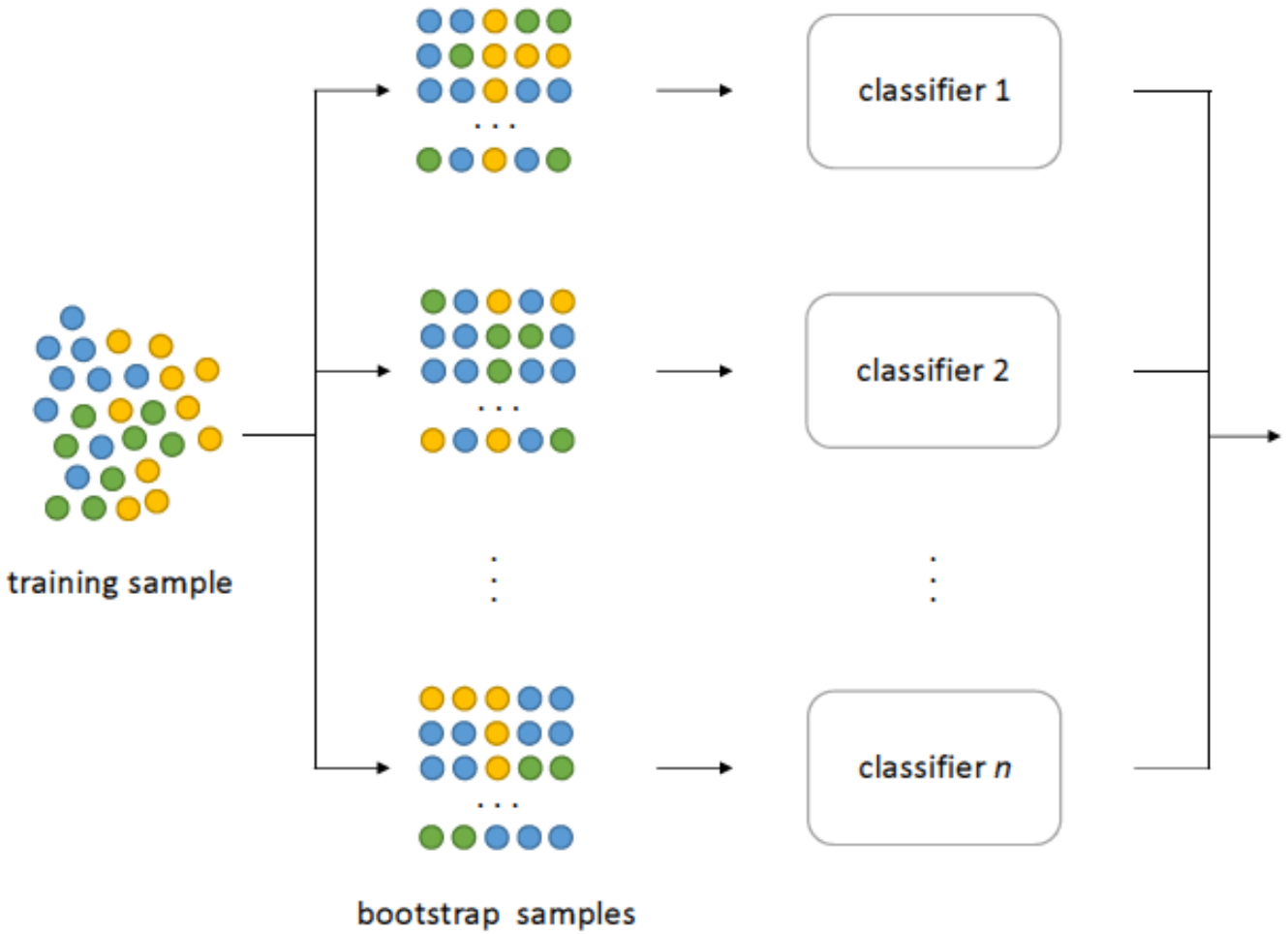
Cost of misclassification of the minority class is higher.

Thus, the region where we should predict 1, or the minority class expands

$$R(i|x) = \sum_{j=1}^{M} P(j|x) \cdot C(i, j),$$



MetaCost



Prediction
Average probability of majority vote

Bayes optimal prediction

Adding the misclassification cost

$$R(i|x) = \sum_{j=1}^{M} P(j|x) \cdot C(i, j),$$

https://www.researchgate.net/figure/The-bagging-approach-Several-classifier-are-trained-on-bootstrap-samples-of-the-training_fig4_322179244

MetaCost - Relabeling

Re-labels the target to the class that minimises the conditional risk

Pseudo code

Inputs:
- S is the training set,
- L is a classification learning algorithm,
- C is a cost matrix,
- m is the number of resamples to generate,
- n is the number of examples in each resample,
- p is *True* iff L produces class probabilities,
- q is *True* iff all resamples are to be used for each example.

Procedure MetaCost (S, L, C, m, n, p, q)

For $i = 1$ to m
  Let S_i be a resample of S with n examples.
  Let M_i = Model produced by applying L to S_i.

For each example x in S
  For each class j
    Let $P(j|x) = \frac{1}{\sum_i 1} \sum_i P(j|x, M_i)$
    Where
      If p then $P(j|x, M_i)$ is produced by M_i
      Else $P(j|x, M_i) = 1$ for the class predicted by M_i for x, and 0 for all others.
      If q then i ranges over all M_i
      Else i ranges over all M_i such that $x \notin S_i$.
  Let x's class = $\text{argmin}_i \sum_j P(j|x)C(i, j)$.

Train In Data

THANK YOU

www.trainindata.com