Train In Data

Wrapping up

The cost in real-world applications

The cost of misclassifying observations of different classes is not the same.

Cost Sensitive Learning (CSL)

- CSL is a that takes (misclassification) costs into account.

- Goal: minimize the total misclassification cost

How do we determine cost?

Cost Matrix provided by Expert

Domain expert to determine real costs
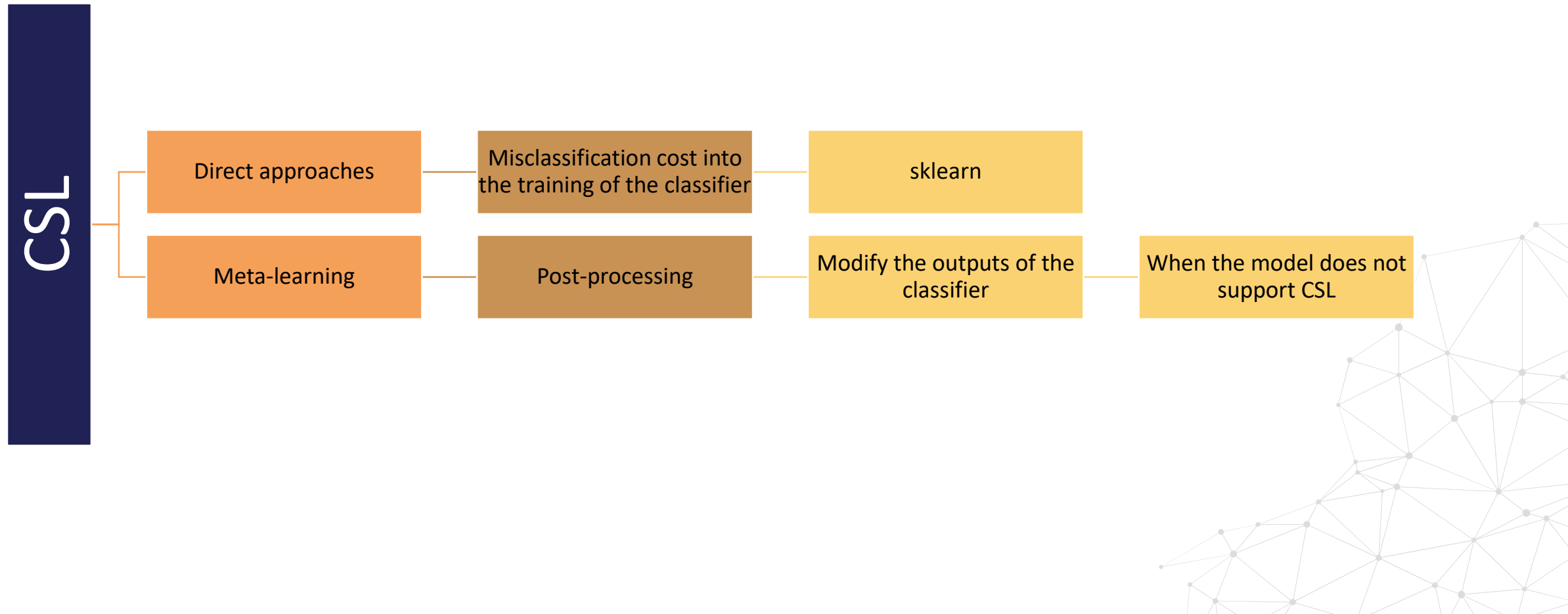
Heuristic

Imbalance Ratio

Optimisation

CSL with sklearn

- Implementing CSL with sklearn is super easy:

  - class_weight parameter

  - sample_weight paramter

- Challenge: determine the right costs

Cost Sensitive Approaches

CSL

- Direct approaches — Misclassification cost into the training of the classifier — sklearn
- Meta-learning — Post-processing — Modify the outputs of the classifier — When the model does not support CSL

Train In Data

THANK YOU

www.trainindata.com