# Introduction

1. Classification algorithms

* Most classification algorithms:
  + Minimize the error rate 🡪 % of incorrect predictions of classes
  + Seek to maximize accuracy
  + Assume that all misclassification errors cost equally

1. The cost in real-world applications

* The cost of misclassifying observations of different classes is not the same
* Misclassification a sick person as healthy is more costly than otherwise
* The patient is at risk if not treated early / properly
* Misclassifying a fraud claim costs more than wrongly classifying a legitimate claim as fraudulent

1. Cost sensitive learning

* CSL is a type of learning that takes (misclassification) costs into account
* Goal: minimize the total misclassification cost
* CSL treats different misclassification differently

1. Cost matrix

* : cost of assigning an observation of class to class
* and : cost of correct classification, usually 0
* and : cost of FN and FP, respectively, usually 1

A screenshot of a computer

AI-generated content may be incorrect.

* Standard ML models use a 0-1 loss function, which assigns a cost of 0 to a correctly classified observation and cost 1 to an incorrectly classified one
* CSL applies different costs to different classification errors

A screenshot of a graph

AI-generated content may be incorrect.

# Types of Cost

1. Misclassification cost

* : cost of assigning an observation of class to class

1. Constant error cost

* Cost of correct classification, usually 0
* Cost of FN and FP, usually 1

1. Questions to consider

* How do we define cost?
* Is cost constant?
* Is cost the same for all observations?
* What does cost depend on?
* Should I care about cost at all?

1. Conditional cost

* Cost of a misclassification error is conditional on the circumstances
* Fraud: the cost of missing a fraudulent application depends on the money involved in the application
* Medical diagnosis: the cost of misclassifying a patient depends on the patient and the disease
* Sensor defect detection: different cost of missing a defect if we have 1 month till the effect occurs or if it is happening now

1. Potential solutions

* Expand the classification target 🡪 different cost to each class
* Healthy / sick and young / sick and elderly
* Defect now / defect in a week / defect in a month / no defect
* Cost of test or feature
* In finance: cost of acquiring variables from 3rd parties
* In medicine: cost of carrying out the tests
  + Money
  + Stress for the patient / side effects
  + Etc.
* Cost of teacher or intervention
* In finance: cost of fraud investigation
* In medicine: cost of a professional
* Pipeline spillage: cost of professional repair
* Computational cost
* Data storage
* Time
* Money
* CO2 for the environment
* Data cost – associated with acquiring the data, in particular the rare class
* Cost of 3rd party labeling
* Cost of in-house labeling
* Cost of buying data from 3rd parties
* Human-computer cost – associated with acquiring the data and building the models
* Data analysts
* Data engineers
* Data scientists
* ML engineers / software developers
* Domain experts
* **Cost is complex. More than 1 player at either side of the scale**

# Obtaining the cost

1. How do we determine cost?

* The effectiveness of CSL relies on supplied cost matrix
* Low cost will not find the proper classification boundary
* High cost may impair generalization
* Ways to obtain cost
* Cost matrix provided by domain experts
* Heuristic
  + Imbalance ratio
  + Optimization
* Factors that influence the ability of a classifier to identify rare events
* Small sample size
* Class separability
* Within-class subclusters

# Cost sensitive approaches

1. Direct approaches

* Introduce misclassification cost into the training of the classifier

1. Meta-learning approaches

* Pre-processing – under/oversampling
* Post-processing – modify the outputs of the classifier

# Misclassification cost in logistic regression

1. Logistic regression – cost function (see slides)
2. Weighted logistic regression - cost function

# Misclassification cost in decision trees

1. Decision trees

* : impurity function
  + Gini:
  + Entropy:
  + Misclassification:
* Weighted decision trees – multiply impurity function by the weight of each classes
* Tree ensembles – same logic
* : proportion of observations of class at each node

# CSL with scikit-learn

1. Misclassification cost as part of training

* Defining parameter for those estimators that allow it, when instantiating the estimator
* Passing a vector with the weights for every single observation when fitting an estimator

1. Parameters

* : can take ‘balanced’ as argument, in which case it will use the balance ratio as weight. Alternatively, it can take a dictionary with {class: penalty} pairs (e.g., {0:1, 1:10} 🡪 misclassification of observations in class 1 are penalized 10 times more than misclassification of class 0)
* : a vector of the same length as y, containing the weight or penalty for each individual observation. 🡪 more flexible

1. Note

* If you use both parameters, the final penalty will be the combination of the 2

1. Classifiers that support