# Classification Algorithms

Most classification algorithms:

- minimize the error rate ➔ the % of incorrect predictions of classes

- seek to maximise the accuracy

- assume that all misclassification errors cost equally

# The cost in real-world applications

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

- Misclassifying a sick person as healthy is more costly than otherwise

  ➔ the patients is at risk if not treated early / properly

- Misclassifying a fraudulent claim costs more (money) than wrongly classifying as fraudulent a legitimate claim

# Cost Sensitive Learning (CSL)

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

- Goal: minimize the total misclassification cost

error ✕ rate

- CSL treats different misclassification differently.

# Cost Sensitive Learning (CSL)

| Cost - insensitive |
| --- |
| - Minimise error rate |
| - Same cost to all misclassification |

| Cost - sensitive |
| --- |
| - Minimise Cost |
| - Different misclassification costs |

Train In Data

# Cost Matrix

C(i,j) ➔ cost of assigning an observation of class j to class i

C(0,0) and C(1,1) ➔ Cost of correct classification, usually 0

C(0,1) and C(1,0) ➔ Cost of FN and FP, respectively, usually 1

| | Real Negative | Real Positive |
|---|---|---|
| Predicted Negative | C(0,0)     TN | C(0,1)     FN |
| Predicted Positive | C(1,0)     FP | C(1,1)     TP |

# Cost Matrix

Standard machine learning 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.

| | Real Negative | Real Positive |
|---|---|---|
| Predicted Negative | C(0,0)         = 0 | C(0,1)        = 1 |
| Predicted Positive | C(1,0)         = 1 | C(1,1)       = 0 |

# Cost Matrix

Cost- sensitive learning applies different costs to different classification errors.

|  | Real Negative | Real Positive |
|---|---|---|
| Predicted Negative | C(0,0)          = 0 | C(0,1)      = 100 |
| Predicted Positive | C(1,0)          = 1 | C(1,1)      = 0 |

# Next ...

- Obtaining the cost

- Types of Cost

- Cost-sensitive approaches