CS456: Machine Learning

Supervised learning: classification

Jakramate Bootkrajang

Department of Computer Science
Chiang Mai University

Objectives

- Understand the basic concepts of data classification
- Able to tell the difference(s) between regression and classification
- Get a big picture of classification models categorisation

Outlines

- Motivation
- Definition
- Classification vs Regression
- Linear vs non-linear machine
- Example of classification tasks
- Major challenges in classification

Motivation

- Suppose there is a fish canning factory which processes two types of fish Sardine and Trout

- Currently, human labours are needed to process the fish.

Motivation

- It might be more cost effective if we have a robotic arm which can 'automatically' separate sardine and trout.

(a) sardine

(b) trout

How do we program such a robot ?

Classical approach: encode everything we know about the fish in our code

```
if 10 < length < 30 and 1 < weight < 5:
    print("Sardine")
else:
    print("Trout")
```

Problems ?

Problems with classical approach

- Our knowledge is incomplete (low accuracy)
- Our robot program might end up being super complicated
- Our robot is not adaptive
  - In the future sardine, in general, might get smaller

Alternative approach

- Let's extract some physical characteristics from sardine and trout
  - e.g., length, weight, colour, existence of body patterns, etc
- And find relationship between physical characteristics and fish types
- Mathematically, find a function which maps from physical characteristics (features) to fish category (label)
- This is a machine learning based classification approach

Definition

Definition

Given a set of features/label pairs $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$, where $\mathbf{x}_i \in R^m$ and $y_i \in \{1, \ldots, K\}$, a classification task is a task of **inferring a real-valued function** $f(\mathbf{x})$ which maps features to its corresponding class label with **high accuracy**.

Definition

- $f()$ is called a classifier
- Each classification model has different functional form $f()$ from the others
- Adjusting function parameters is accounted for the learning part [1]

[1] There exist classifiers which do not assume function form such as k-NN, Decision Tree. Such classifiers are called non-parametric models (won't be covered in this class)

The goal of learning

- To minimise classifier error
- Classification error is defined as

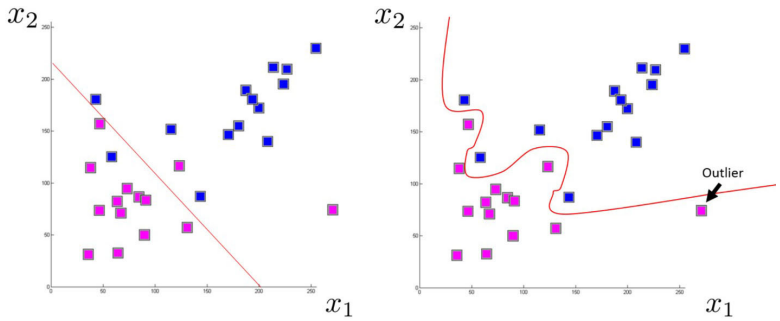
$$\epsilon = \sum_{i=1}^{N} \delta(y_i \neq f(\mathbf{x}_i)) \tag{1}$$

- Delta function, $\delta(q)$ returns 1 if the logical expression q is true, and returns 0 otherwise
- Note that accuracy is the inverse of error

Classification vs Regression

- Regression
  - For $(\mathbf{x}, y)$, the output y is called the target and $y \in \mathcal{R}$
- Classification
  - For $(\mathbf{x}, y)$, the output y is called the class label and $y \in \mathcal{I}$
  - Output from regression model can be transformed into label prediction by thresholding
- Both belong to supervised learning paradigm

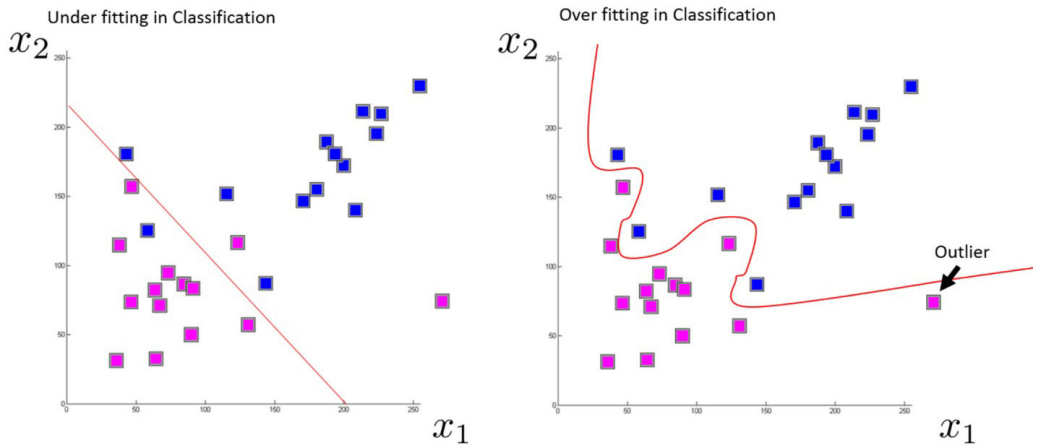
Types of classifier by functional form



- Linear
  - Assumes data is separable using linear decision boundary
  - Least Square Classifier, Logistic regression, Support Vector Machine, Neural networks without hidden layer
- Non-linear
  - Assumes data is separable using non-linear decision boundary
  - SVM (with kernel), Neural networks, Ensemble of classifiers (boosting), Quadratic Discriminant Analysis

Pros/Cons of linear vs nonlinear

- Linear
  - Works well in linearly separable data
  - May underfit in complex cases
- Non-linear
  - Works well in linearly inseparable data
  - May overfit in simpler cases



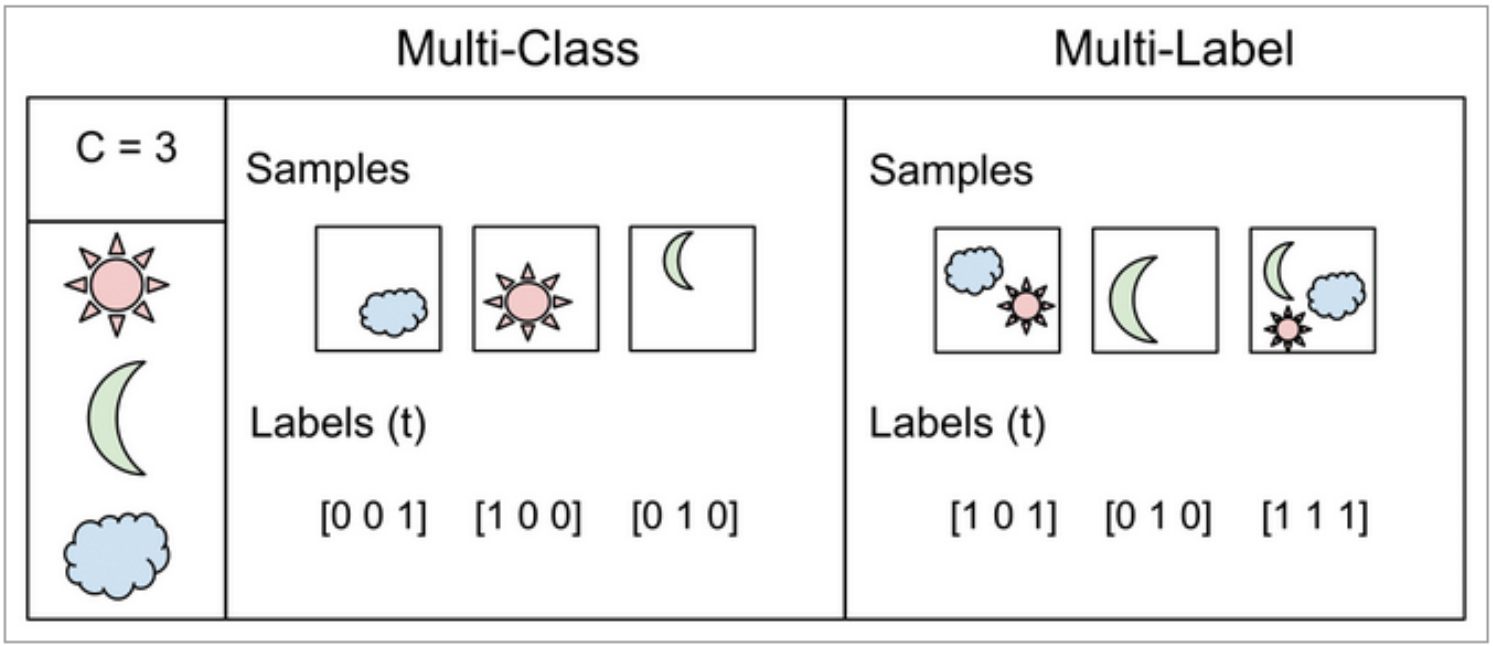
Classification tasks/applications

- Binary classification
- Multiclass classification
- Multi-label classification
- One-class classification
- Openset classification

Multi-class vs multi-label

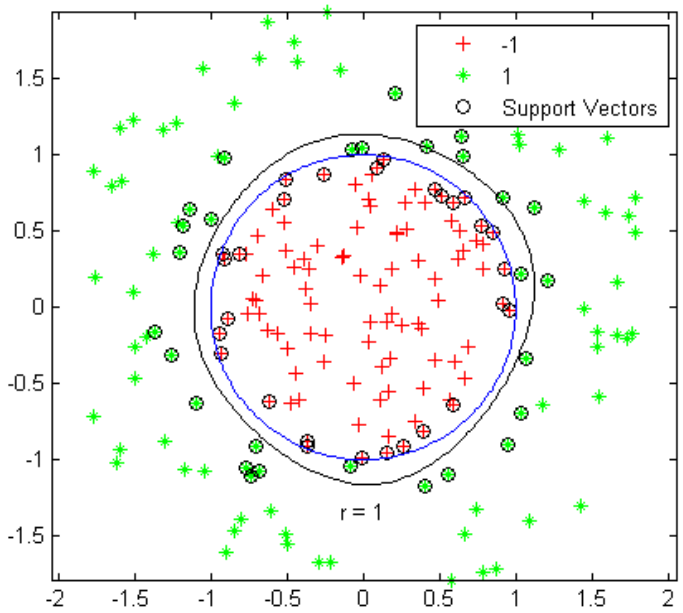
Multi-class: there are $K > 2$ data classes and one input can belong to only one class

Multi-label: there are $K > 2$ data classes and one input can belong to more than one class



Application: medical diagnosis etc.

One class classification



Application: person identification, object detection

Negative examples are often countless: impractical to use binary classification

Open set classification

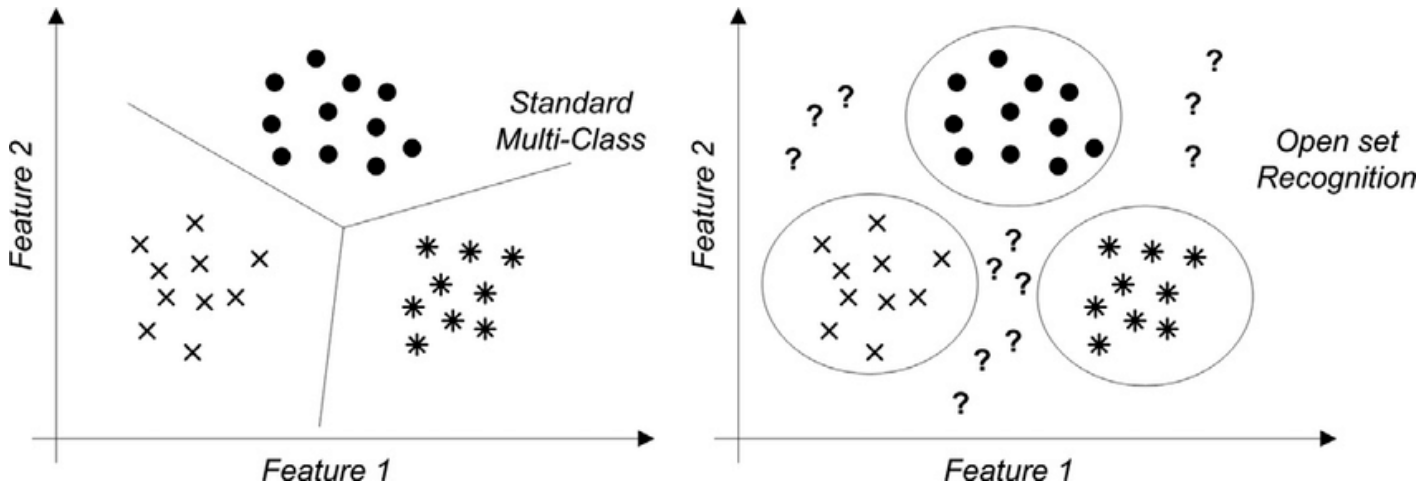


Figure credit [2]

[2] Lazzeratti et al.: A new approach for event classification and novelty detection in power distribution networks

Major challenges

- Dataset size vs model complexity (e.g., the number of parameters)
- Data dimensionality
- Incorrect features / labels
- Class imbalance
- Data distribution is dynamic

Objectives: revisited

- Understand the basic concepts of data classification
- Able to tell the difference(s) between regression and classification
- Get a big picture of classification models categorisation

Lastly

Questions please ..