Applied Text Analytics & Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Classification Introduction

These slides are inspired based on slides from Mahdi Roozbahani, Le Song, Chao Zhang, Yaser Abu-Mostafa, Andrew Zisserman, Polo Chau.

Learning Objectives

In this lesson, you will learn about general concept of linear classification

- Supervised Learning
- Regression

Learning Objectives

In this lesson, you will learn about general concept of linear classification

- Supervised Learning
- Regression

Supervised Learning: Overview

Functions $\mathcal{F}$

$f : \mathcal{X} \to \mathcal{Y}$

Training data

$\{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}$

LEARNING

find $\hat{f} \in \mathcal{F}$
s.t. $y_i \approx \hat{f}(x_i)$

Learning machine

New data

PREDICTION

$y = \hat{f}(x)$ ⇐ x

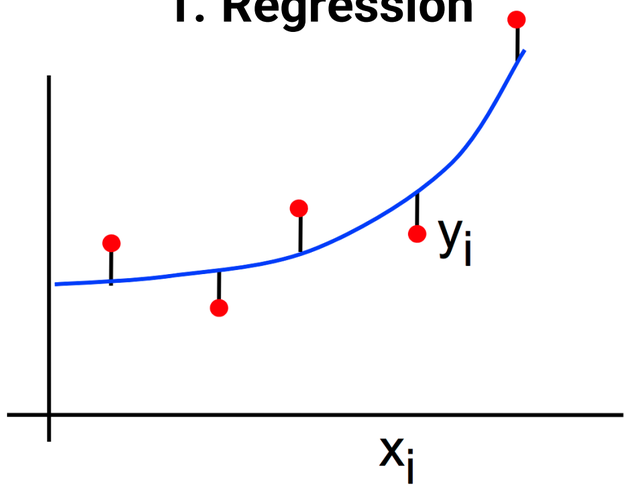
Supervised Learning: Two Types of Tasks

Given: training data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$

Learn: a function $f(\mathbf{x}) : y = f(\mathbf{x})$

When y is continuous:

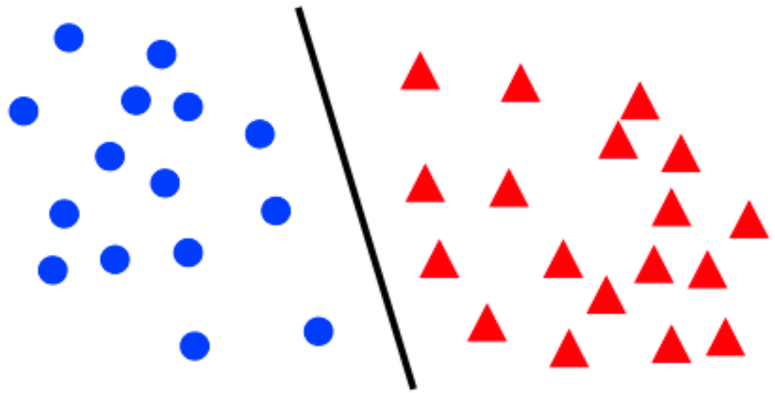
1. Regression



Curve fitting

When y is discrete:

2. Classification



Class estimation

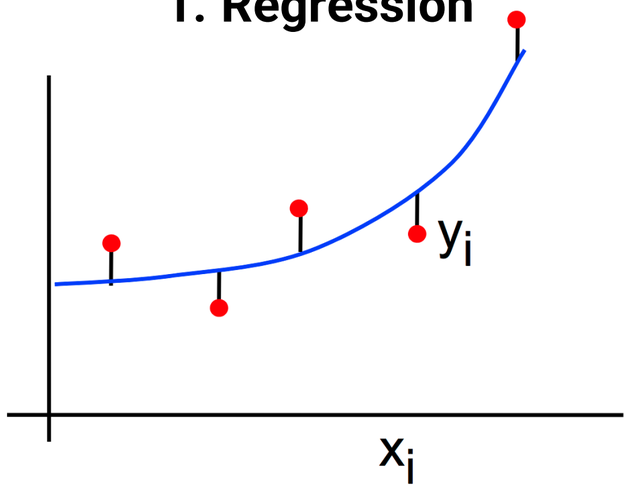
Supervised Learning: Two Types of Tasks

Given: training data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$

Learn: a function $f(\mathbf{x}) : y = f(\mathbf{x})$

When y is continuous:

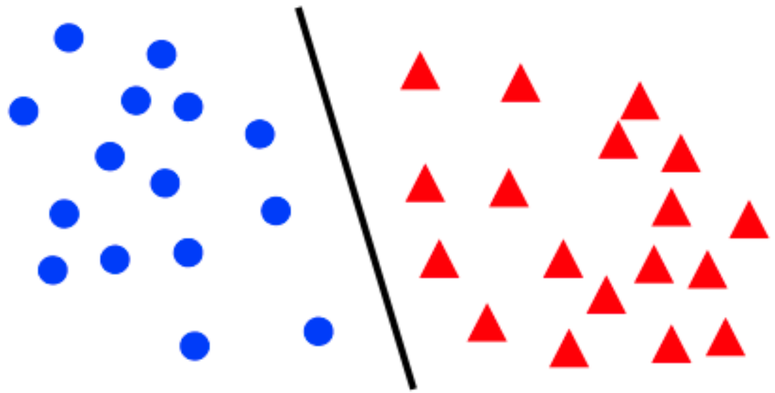
1. Regression



Curve fitting

When y is discrete:

2. Classification



Class estimation

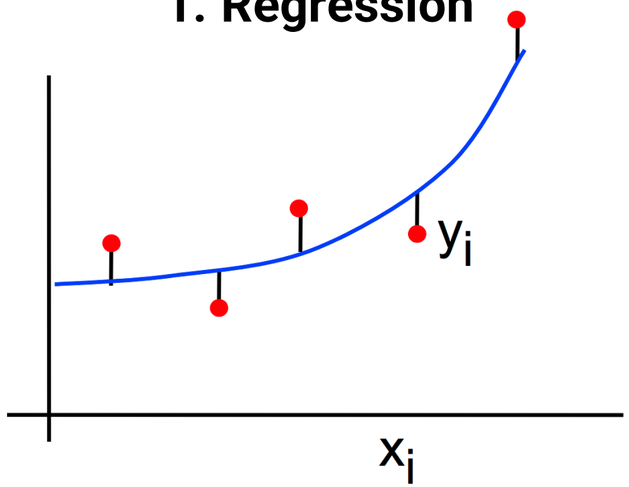
Supervised Learning: Two Types of Tasks

Given: training data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$

Learn: a function $f(\mathbf{x}) : y = f(\mathbf{x})$

When y is continuous:

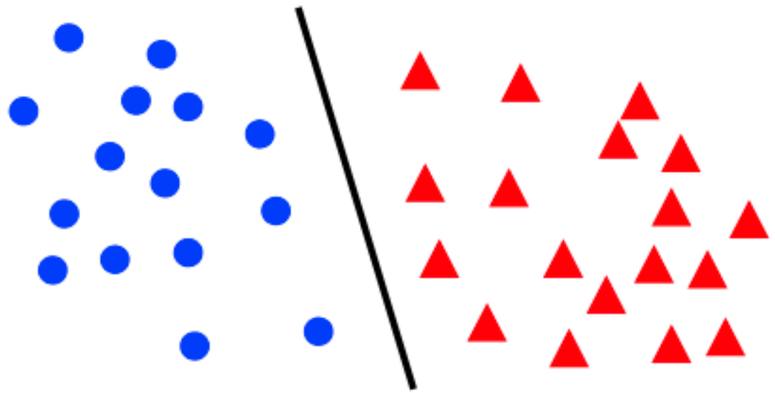
1. Regression



Curve fitting

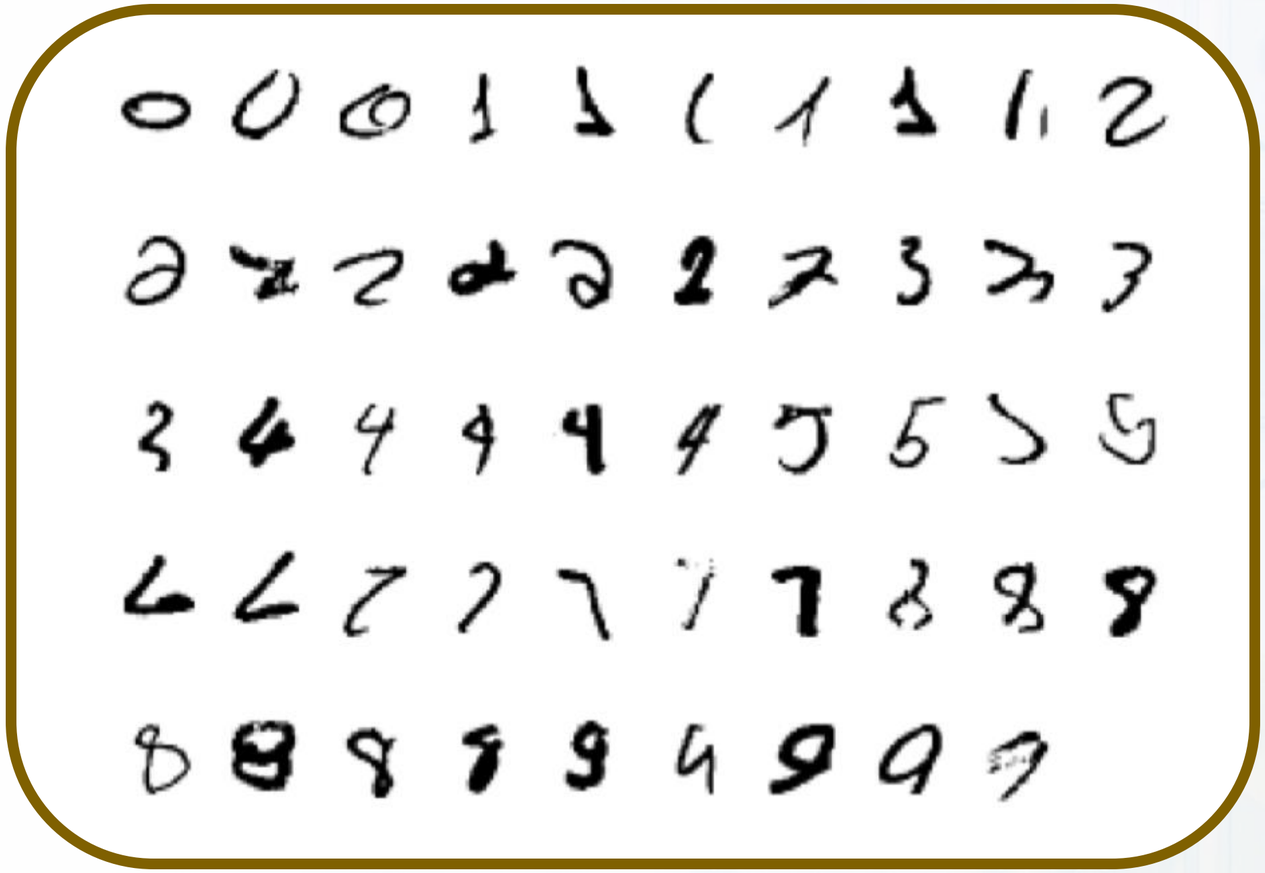
When y is discrete:

2. Classification

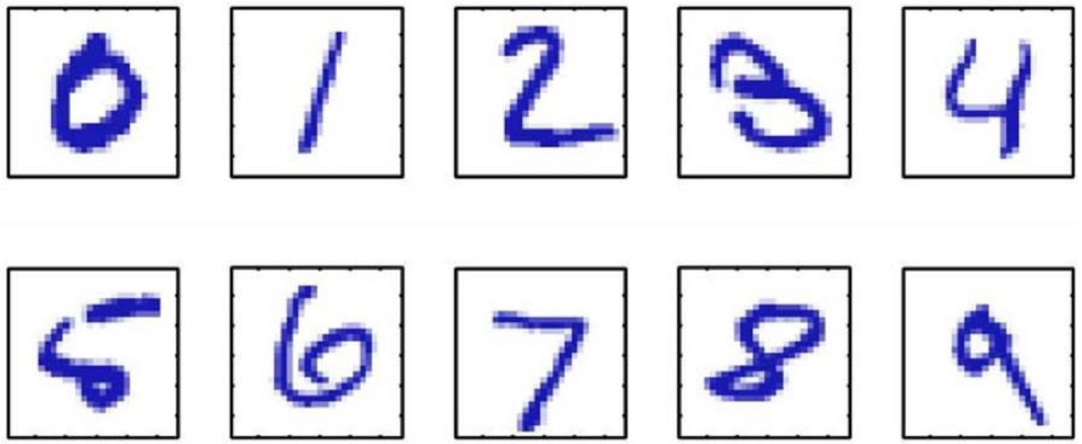


Class estimation

Classification Ex 1: Handwritten Digit Recognition



Classification Ex 1: Handwritten Digit Recognition



Images are 28 x 28 pixels

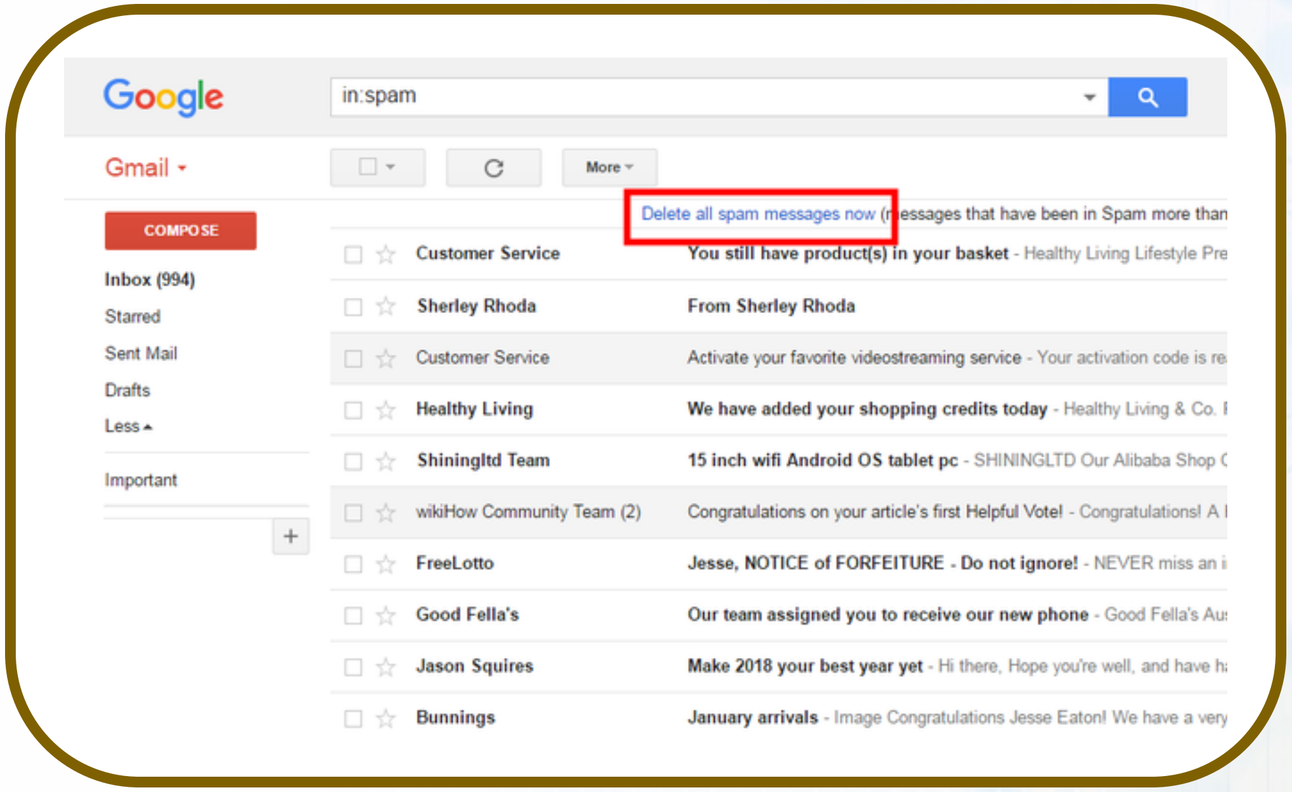
A classification problem

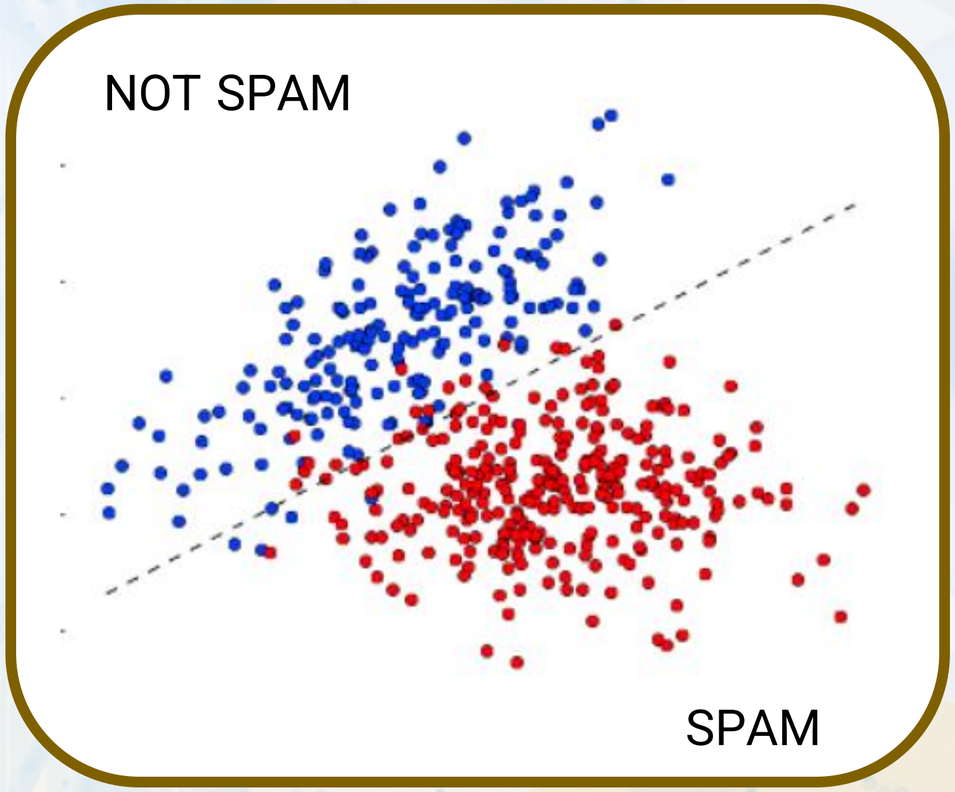
Represent input image as a vector $\mathrm{x} \in \mathbb{R}^{784}$

Learn a classifier $f(\mathrm{x})$ such that

$$f: \mathrm{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Classification Ex 2: Spam Detection





Regression Ex 1: Apartment Rent Prediction

- Suppose you are to move to Atlanta
- And you want to find the **most reasonably priced** apartment satisfying your **needs:**
    - Square-ft., # of bedroom, distance to campus…

Living Area (ft^2)	# Bedroom	Rent ($)
230	1	600
506	2	1000
433	2	1100
109	1	500
…		
150	1	?
270	1.5	?

Regression Ex 2: Stock Price Prediction

Image taken from: Chen, J., 2021. *Volume Price Trend Indicator (VPT)*. [online] Investopedia. Available at: <https://www.investopedia.com/terms/v/vptindicator.asp> [Accessed 18 May 2022].

Summary

- Supervised learning paradigm
- Classification vs Regression