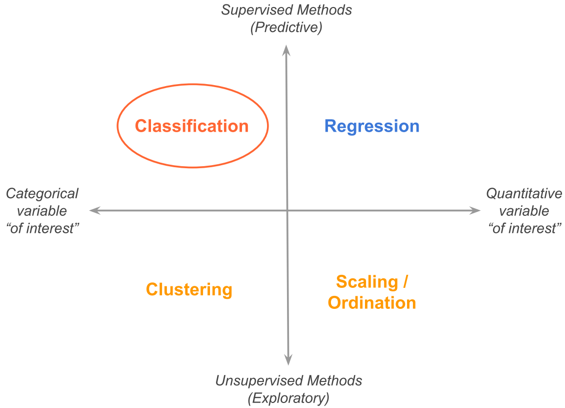
Classification

Ngoc Hoang Luong

University of Information Technology (UIT), VNU-HCM

May 27, 2024

Introduction

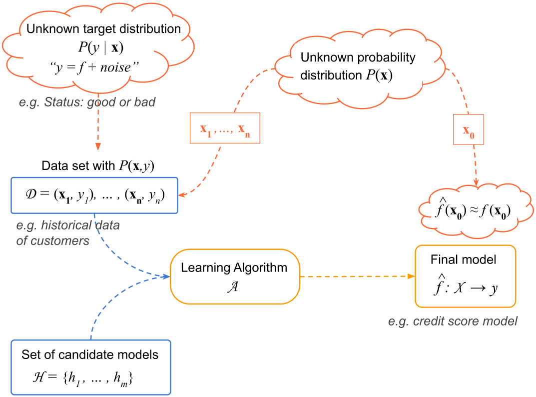


- The goal in classification is to take an input vector $\mathbf{x}$ and to assign it into one of K discrete classes or groups C_k where $k = 1, 2, \ldots, K$.
- The classes are assumed to be disjoint, i.e., each input is assigned to one and only one class.

Classification - Example

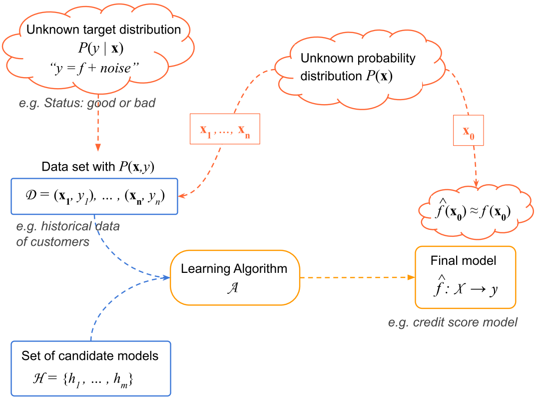
- Consider a credit application with p predictors $X = [X_1, \ldots, X_p]$: Age, Salary, Residential Status, Marital Status, Debt, etc.
- A **credit score** is computed for each application to relate how like each applicant can pay the debt.
- Customers are divided into two classes: *good* and *bad*:
  - Good customers are those that payed their loan back.
  - Bad customers are those that defaulted on their loan

Classification - Supervised Learning Diagram



- Joint distribution of data: $P(\mathbf{x}, y)$
- Conditional distribution of target, given inputs $P(y \mid \mathbf{x})$
- Marginal distribution of inputs $P(\mathbf{x})$

Classification - Supervised Learning Diagram

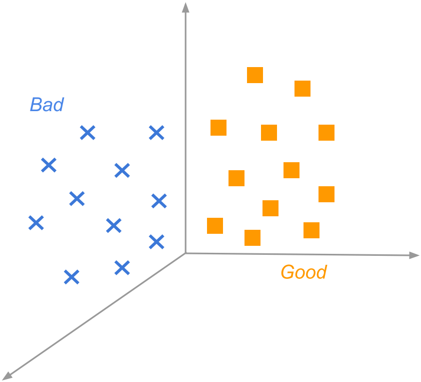


- The idea in classification problems is: Given a customer's attributes $X = \mathbf{x}$, to which class y we should assign this customer?
- We would like to know what is **the conditional probability**:

$$P(y \mid X = \mathbf{x})$$

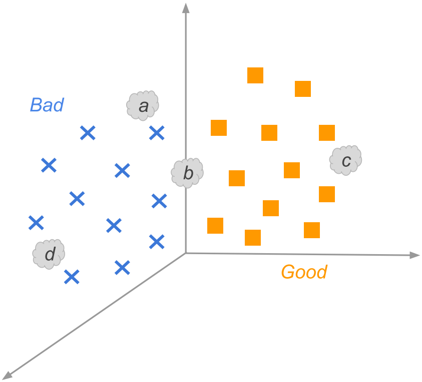
Classification - Example

- Suppose we have n individuals in a p-dimensional space.
- Suppose each class of customers forms its own cloud: the good customers, the bad customers.



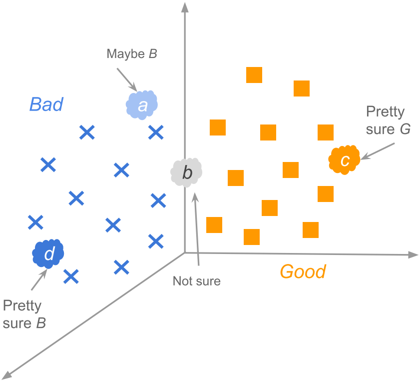
Classification - Example

- Now, assume there are four individuals a, b, c, d that we want to predict their classes.
- We want to have a mechanism or **rule** to classify observations.



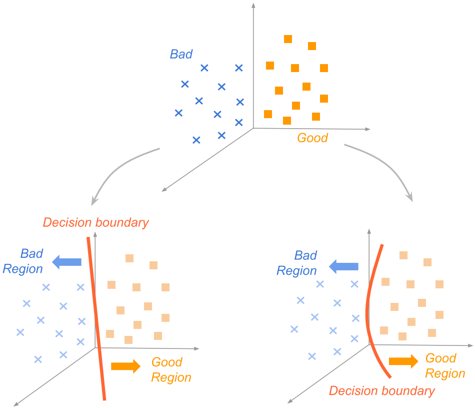
Classification - Example

- Customer a could be assigned to class bad.
- Customer d could also be assigned to class bad with high confidence.
- Customer c could be assigned with high confidence to class good.
- We could be uncertain to which class customer b belongs.

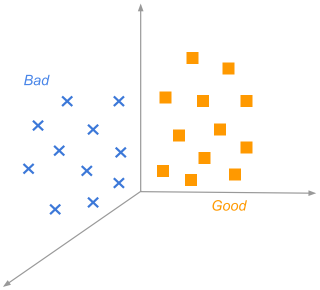


Classification - Example

- Classification rules allow us to divide the input space into regions $\mathcal{R}_k$ called **decision regions** (one for each class).
- The boundaries between decision regions establish the decision boundaries or decision surfaces.



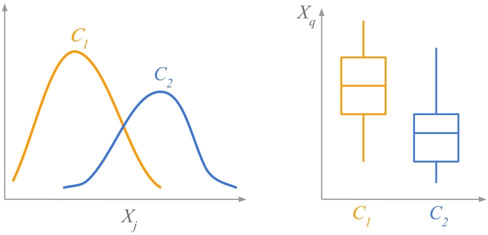
Classification - Two-class Example



- We have customers belonging to one of two classes C_1 = good and C_2 = bad.
- We can first investigate how X values vary according to a given class C_k - the class-conditional distribution:

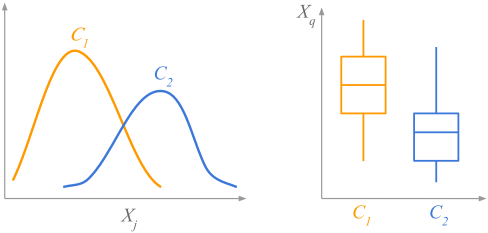
$$P(X = \mathbf{x} \mid y = k)$$

Classification - Exploring Conditional Distributions



- How does $X_j \mid y = 1$ compare with $X_j \mid y = 2$?
- How does $X_q \mid y = 1$ compare with $X_q \mid y = 2$?
- From data, we can have descriptive information about $X \mid y = k$. We calculate summary statistics, compare visual displays of these distributions.

Classification - Exploring Conditional Distributions



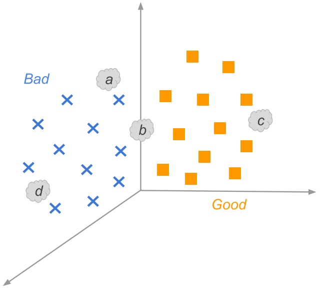
- If we have the class-conditional distribution $P(X \mid y = k)$, we can compute:

$$P(X = \mathbf{x} \mid \text{Good}) = \frac{\text{applicant is Good and has attributes } \mathbf{x}}{\text{applicant is Good}}$$

or

$$P(X = \mathbf{x} \mid \text{Bad}) = \frac{\text{applicant is Bad and has attributes } \mathbf{x}}{\text{applicant is Bad}}$$

Classification - Conditional Probability



- However, we are actually interested in the conditional probability $P(y = k \mid X = \mathbf{x})$, we can compute:

$$P(\text{Good} \mid X = \mathbf{x}) = \frac{\text{applicant is Good and has attributes } \mathbf{x}}{\text{applicant has attributes } \mathbf{x}}$$

or

$$P(\text{Bad} \mid X = \mathbf{x}) = \frac{\text{applicant is Bad and has attributes } \mathbf{x}}{\text{applicant has attributes } \mathbf{x}}$$

Bayes' Rule Reminder

- We have the conditional probabilities:

$$P(X = \mathbf{x} \mid y = k) = \frac{P(y = k, X = \mathbf{x})}{P(y = k)}$$

  and

$$P(y = k \mid X = \mathbf{x}) = \frac{P(y = k, X = \mathbf{x})}{P(X = \mathbf{x})}$$

- We have the joint probability:

$$\begin{aligned} P(X = \mathbf{x}, y = k) &= P(y = k \mid X = \mathbf{x})P(X = \mathbf{x}) \\ &= P(X = \mathbf{x} \mid y = k)P(y = k) \end{aligned}$$

- Thus, we have:

$$P(y = k \mid X = \mathbf{x}) = \frac{P(X = \mathbf{x} \mid y = k)P(y = k)}{P(X = \mathbf{x})}$$

Bayes' Rule Reminder

$$P(y = k \mid X = \mathbf{x}) = \frac{P(X = \mathbf{x} \mid y = k)P(y = k)}{P(X = \mathbf{x})}$$

where the marginal probability $P(X = \mathbf{x})$ can be computed with the total probability formula:

$$P(X = \mathbf{x}) = \sum_k P(X = \mathbf{x} \mid y = k)P(y = k)$$

We can use Bayes' Theorem for classification purpose:

- $P(y = k) = \pi_k$: the prior probability for class k.
- $P(X = \mathbf{x} \mid y = k) = f_k(\mathbf{x})$: the class-conditional density for inputs X in class k.

The **posterior probability** (the conditional probability of the response given the input) is:

$$P(y = k \mid X = \mathbf{x}) = \frac{f_k(\mathbf{x})\pi_k}{\sum_{k=1}^{K} f_k(\mathbf{x})\pi_k}$$

Bayes' Rule Reminder

- The posterior probability:

$$P(y = k \mid X = \mathbf{x}) = \frac{f_k(\mathbf{x})\pi_k}{\sum_{k=1}^{K} f_k(\mathbf{x})\pi_k}$$

- By using Bayes' Theorem, we are modeling the posterior probability $P(y = k \mid X = \mathbf{x})$ in terms of likelihood densities $f_k(\mathbf{x})$ and prior probabilities π_k.

$$\texttt{posterior} = \frac{\texttt{likelihood} \times \texttt{prior}}{\texttt{evidence}}$$

Bayes Classifiers

- In supervised learning, the goal is to find a model $\hat{f}()$ that makes good predictions.
- In a classification setting, we minimize the probability of assigning an individual $\mathbf{x}_i$ to the wrong class.
- We should classify $\mathbf{x}_i$ to the class k that makes $P(y = k \mid X = \mathbf{x})$ as large as possible, i.e., classify $\mathbf{x}_i$ to the most likely class, given its predictors.