Introduction to Machine Learning

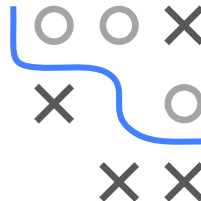## Supervised Classification
In a Nutshell
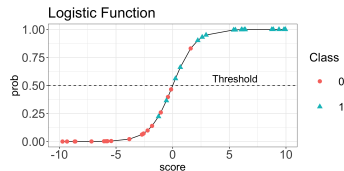
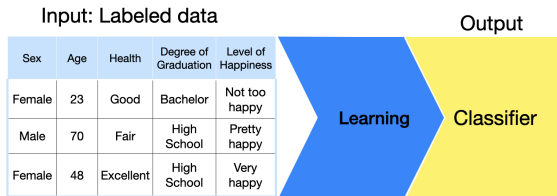Logistic Function

Learning goals

- Understand basic concept of classifiers

- Know concepts of probabilistic and scoring classifier

- Know distinction between discriminant and generative approach

- Understand ideas of logistic regression and Naive Bayes
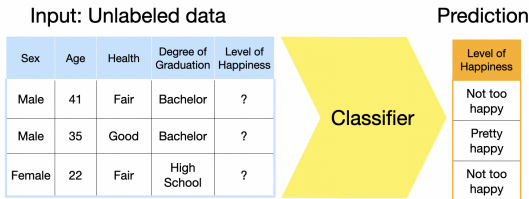
CLASSIFICATION TASKS

- Learn function that assigns categorical class labels to observations
- Each observation belongs to exactly one class
- The task can contain two (binary) or multiple (multi-class) classes


Training

Input: Labeled data

Sex	Age	Health	Degree of Graduation	Level of Happiness
Female	23	Good	Bachelor	Not too happy
Male	70	Fair	High School	Pretty happy
Female	48	Excellent	High School	Very happy

Learning → Classifier

Output

Prediction

Input: Unlabeled data

Sex	Age	Health	Degree of Graduation	Level of Happiness
Male	41	Fair	Bachelor	?
Male	35	Good	Bachelor	?
Female	22	Fair	High School	?

Classifier →

Prediction

Level of Happiness
Not too happy
Pretty happy
Not too happy

BASIC DEFINITIONS

- For every observation a model outputs the probability (probabilistic classifier) or score (scoring classifier) of each class
- In the multi-class case, the class label is usually assigned by choosing the class with the maximum score or probability
- In the binary case, a class label is assigned by choosing the class whose probability or score exceeds a threshold value c
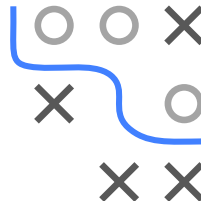
Input: Unlabeled data

Sex	Age	Health	Degree of Graduation	Level of Happiness
Male	41	Fair	Bachelor	?

Class Probabilities

| Probability	Level of Happiness
0.4	Not too happy
0.35	Pretty happy
0.25	Very happy

Assigned Label

Level of Happiness
Not too happy

LOGISTIC REGRESSION

- Logistic regression is a **discriminant approach** for binary classification. It turns scores into probabilities with the logistic function.
- We just need to compute the probability for **one** class (usually class 1).
- If the probability exceeds a threshold value **c** $\Rightarrow$ class 1 is predicted.
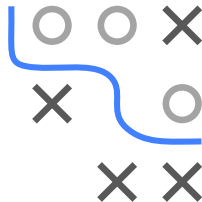

Input: Unlabeled data

X1	X2	Class
2.3	1.7	?

Classifier
$Score = \theta_0 + \theta_1 X_1 + \theta_2 X_2$

Scores

Score	Class
1.2	?

Logistic Function
$\frac{exp(score)}{(1 + exp(score))}$

Class Probabilities

Probability	Class
0.7685	?

Threshold
c = 0.5

Assigned Label
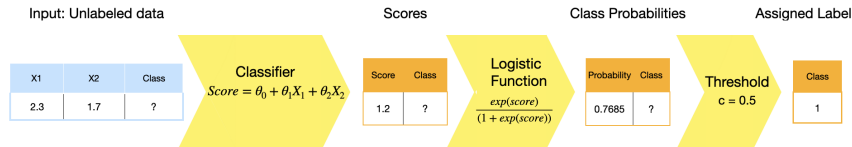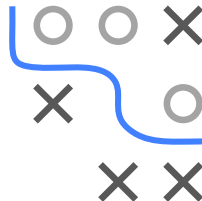
Class
1

The logistic function puts all scores in order along an s-shaped line.

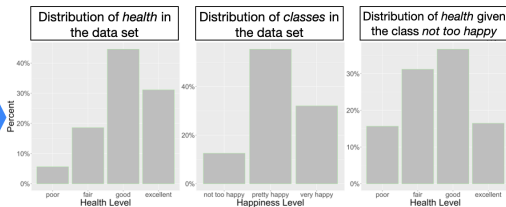Logistic Function
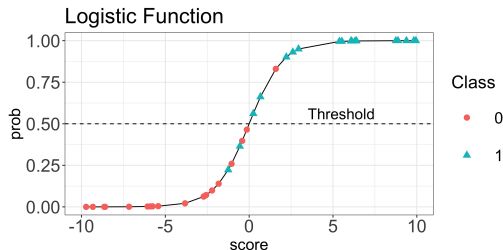
NAIVE BAYES

- Naive Bayes is a **generative multi-class approach**. It computes the class probability for each class based on the training data.

- It considers the data distribution on three different levels:
  - Marginal distributions $\mathbb{P}(X)$ of each feature (in the entire data set)
  - Marginal distribution $\mathbb{P}(Y)$ of classes (in the entire data set)
  - Conditional distributions $\mathbb{P}(X|Y)$ of each feature in each class