

CLASSIFICATION TASKS

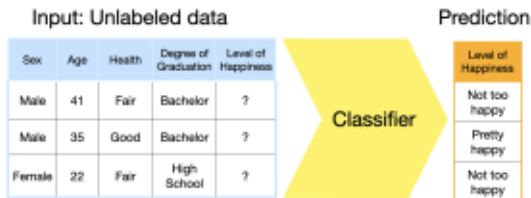
- Learn function that assigns categorical class labels to observations
- Each observation belongs to exactly one class
- The task can contain two (binary) or multiple (multi-class) classes



Training



Prediction



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	?

Classifier

Class Probabilities

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

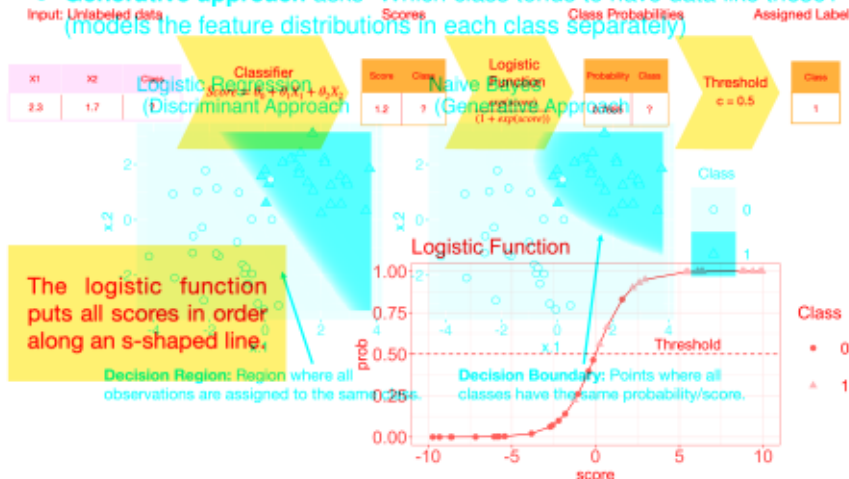
Assigned Label

Level of Happiness
Not too happy

LOGISTIC REGRESSION

Two fundamental approaches exist to construct a classifier:

- **Discriminant approach** asks "What is the best prediction for the class given these data?" (uses loss functions and empirical risk minimization)
- 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.
- **Generative approach** asks "Which class tends to have data like these?" (models the feature distributions in each class separately)



NAIVE BAYES REGRESSION

- Logistic regression is a discriminant approach for binary classification. It turns scores into probabilities with the logistic function.
- Naive Bayes is a **generative multi-class approach**. It computes the class probability for each class based on the training data.
- We just need to compute the probability for **one** class (usually class 1).
- It considers the data distribution on three different levels:
 - Marginal distributions $P(X)$ of each feature (in the entire data set)
 - Marginal distribution $P(Y)$ of classes (in the entire data set)
 - Conditional distributions $P(X|Y)$ of each feature in each class

Input: Unlabeled data

X1	X2	Class
2.3	1.7	?

$$\text{Score} = \theta_0 + \theta_1 X_1 + \theta_2 X_2$$

Score	Class
1.2	?

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

Probability	Class
0.7666	?

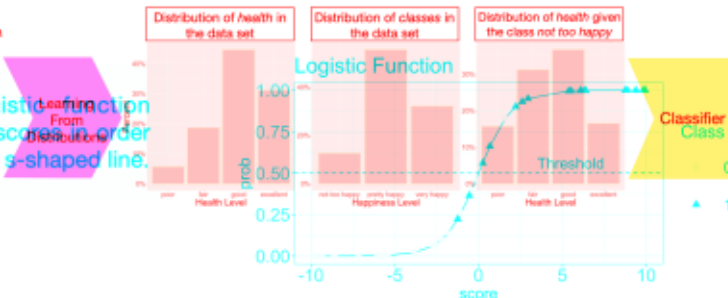
Threshold	Class
c = 0.5	?

Class
1



Input: Labeled data

Health	Level of Happiness
Good	Not too happy
Fair	Pretty happy
OK	Very happy
Excellent	Very happy



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 $P(X)$ of each feature (in the entire data set)
 - Marginal distribution $P(Y)$ of classes (in the entire data set)
 - Conditional distributions $P(X|Y)$ of each feature in each class



Input: Labeled data

Health	Level of Happiness
Good	Not too happy
Fair	Pretty happy
Fair	Fair
Excellent	Very happy

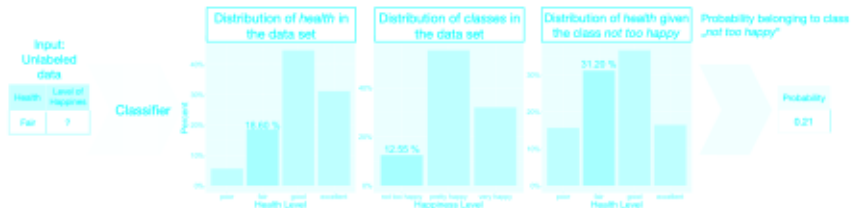
Learning From Distributions



Classifier

NAIVE BAYES

- Example: Class probability of "not too happy" given health = "fair":



Naive Bayes Classifier

$$\frac{P(X = x|Y = k) \times P(Y = k)}{P(X = x)}$$

31.20 % 12.55 %

18.60 %

Class probability
given the data

21.00 %