Hochschule Offenburg
offenburg.university

Machine Learning III

Simple Probabilistic Models

Recall Classification

Supervised Learning: Annotated Training Data

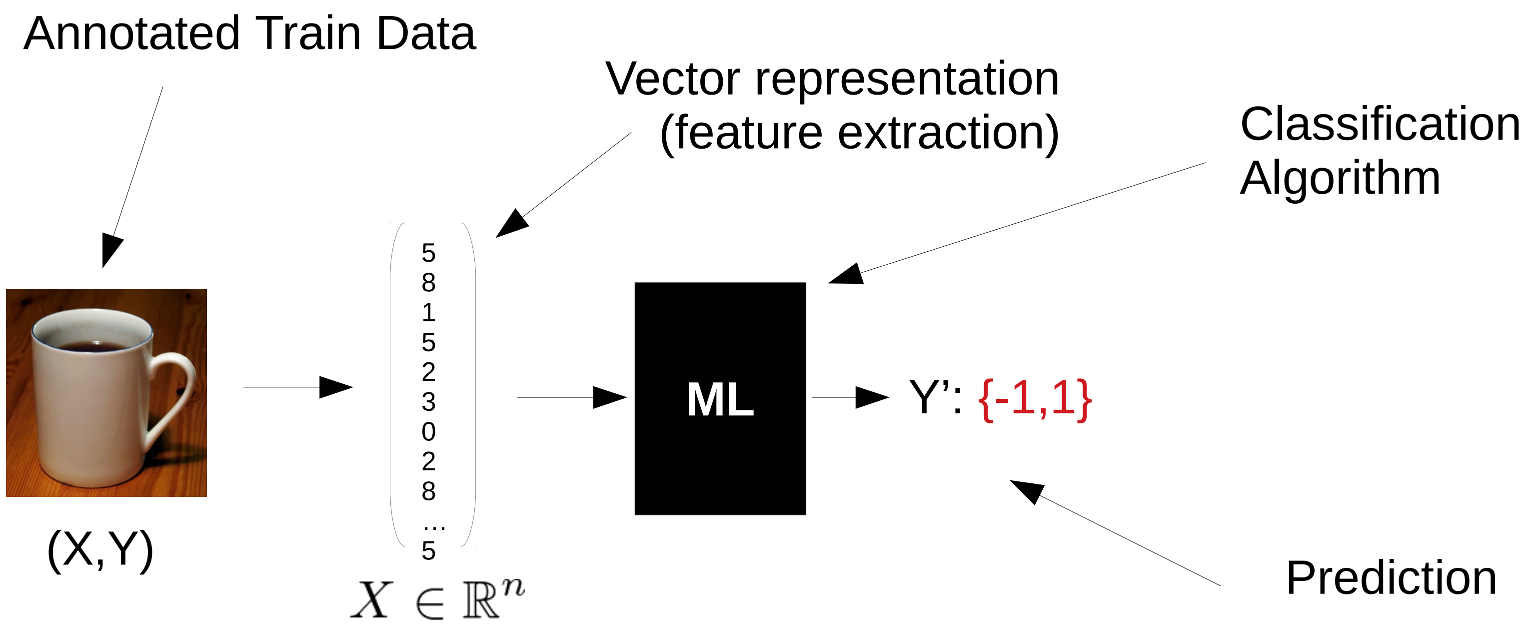
(X,Y)

$$\begin{pmatrix} 5 \\ 8 \\ 1 \\ 5 \\ 2 \\ 3 \\ 0 \\ 2 \\ 8 \\ \dots \\ 5 \end{pmatrix} \rightarrow \textbf{ML} \rightarrow Y': \{-1,1\}$$

$$X \in \mathbb{R}^n$$

Recall Classification

Supervised Learning: Annotated Training Data



Recall Classification

LEARNING: is a optimization problem → Finding the best function separating

$$min(\|f(X,Y),Y'\|)$$

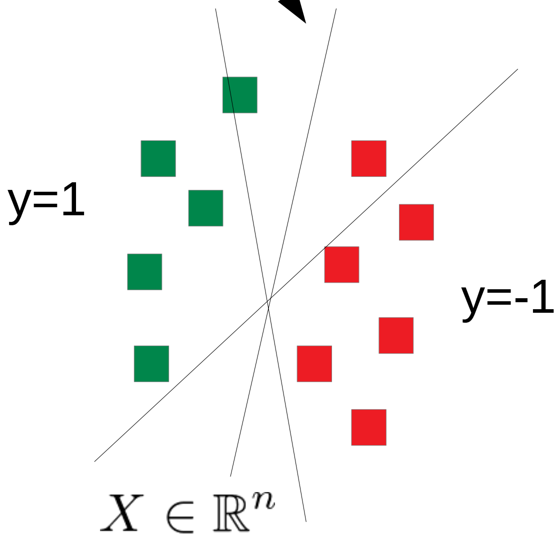
(X,Y)

$\begin{pmatrix} 5 \\ 8 \\ 1 \\ 5 \\ 2 \\ 3 \\ 0 \\ 2 \\ 8 \\ \dots \\ 5 \end{pmatrix}$

$X \in \mathbb{R}^n$

ML

Y': {-1,1}



$X \in \mathbb{R}^n$

A Simple 1D Example

1D Feature Space. All data samples a simple scalar values.

Task: “learn” splitting function from data.

x

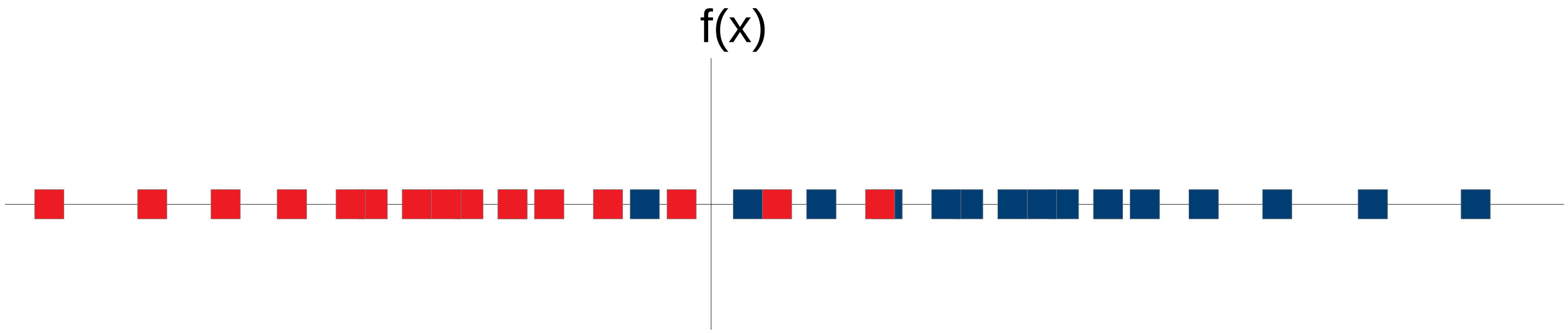
y: color

A Simple 1D Example

1D Feature Space. All data samples a simple scalar values.

Task: “learn” splitting function from data.

Where should we cut?

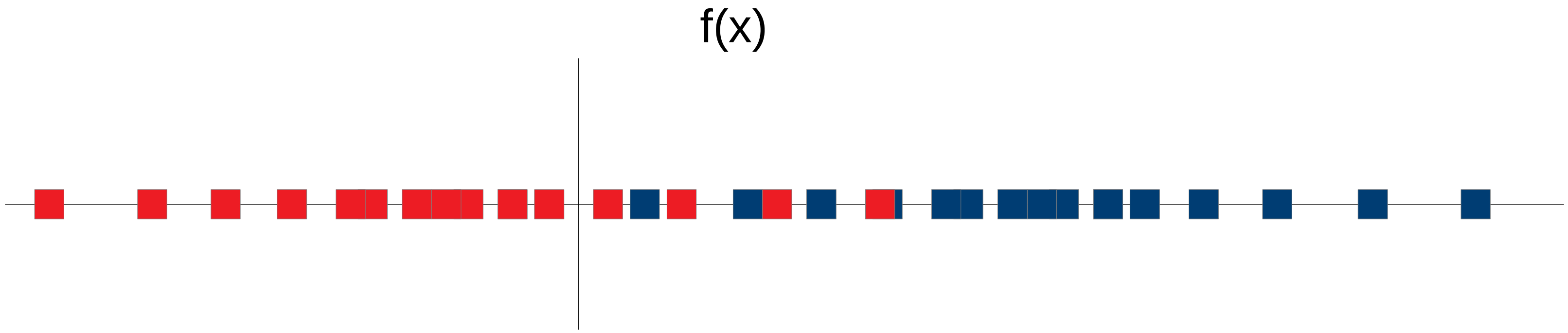


A Simple 1D Example

1D Feature Space. All data samples a simple scalar values.

Task: “learn” splitting function from data.

Where should we cut?

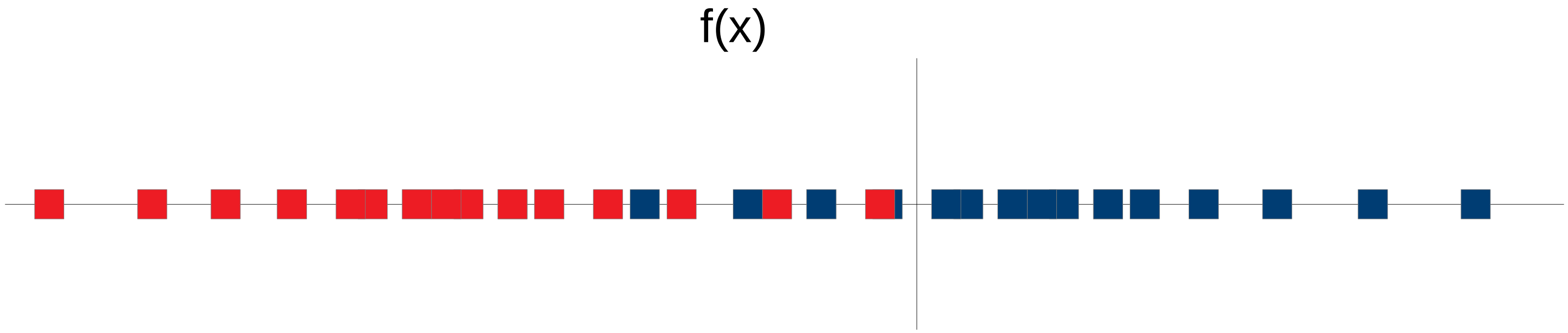


A Simple 1D Example

1D Feature Space. All data samples a simple scalar values.

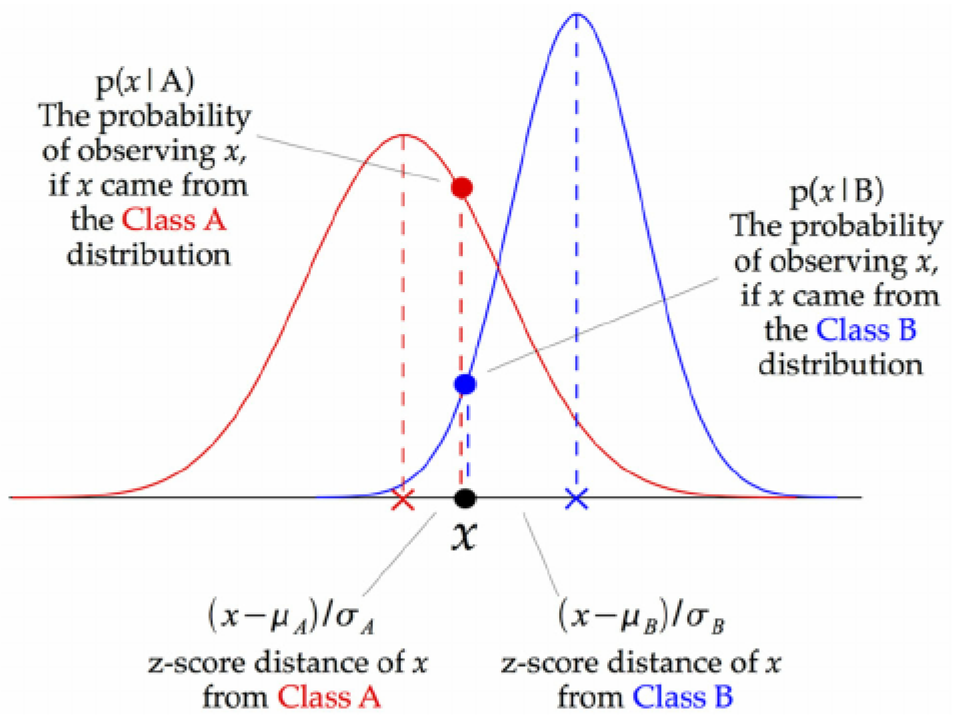
Task: “learn” splitting function from data.

Where should we cut?



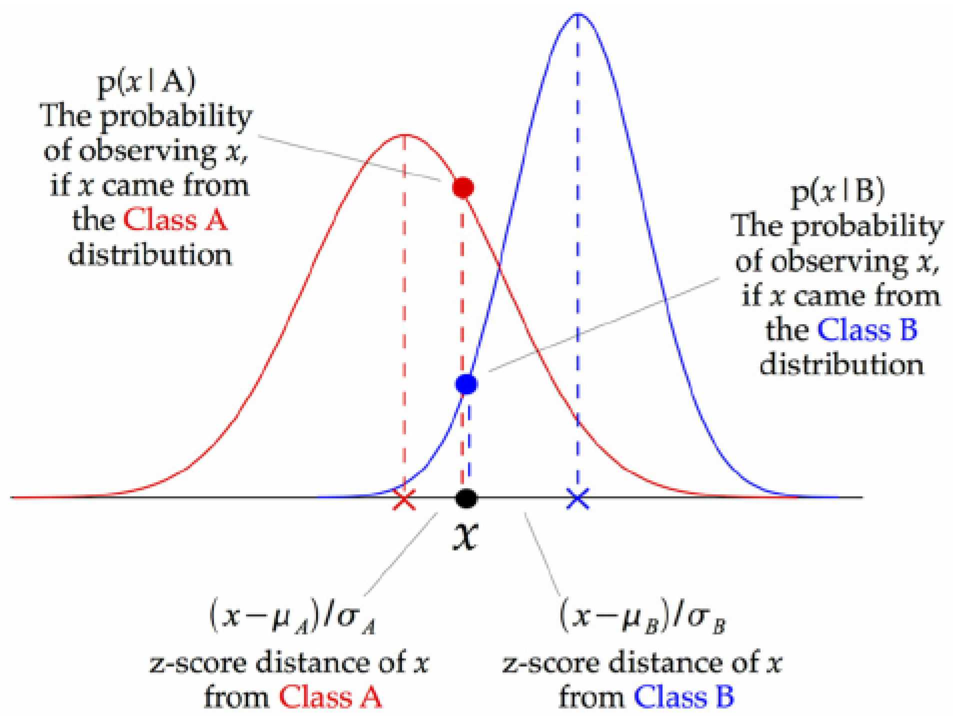
Naive Bayes Model (1D)

Here Gaussian Bayes (other distributions possible!)



Naive Bayes Model (1D)

Here Gaussian Bayes (other distributions possible!)

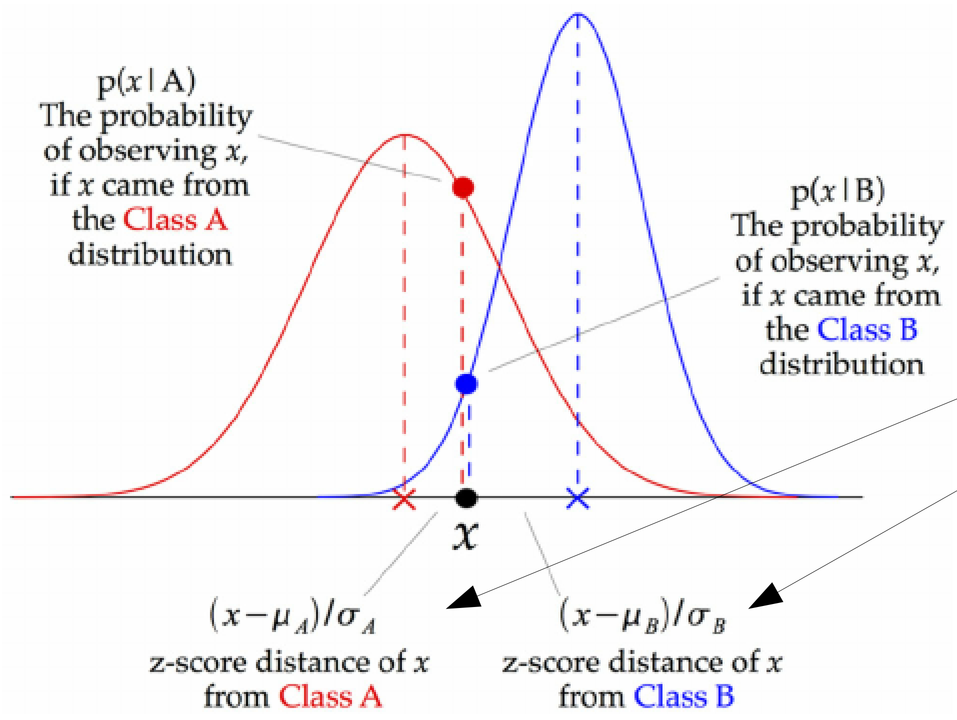


Bayes Rule:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

Naive Bayes Model (1D)

Here Gaussian Bayes (other distributions possible!)



Bayes Rule:
Compute for all classes

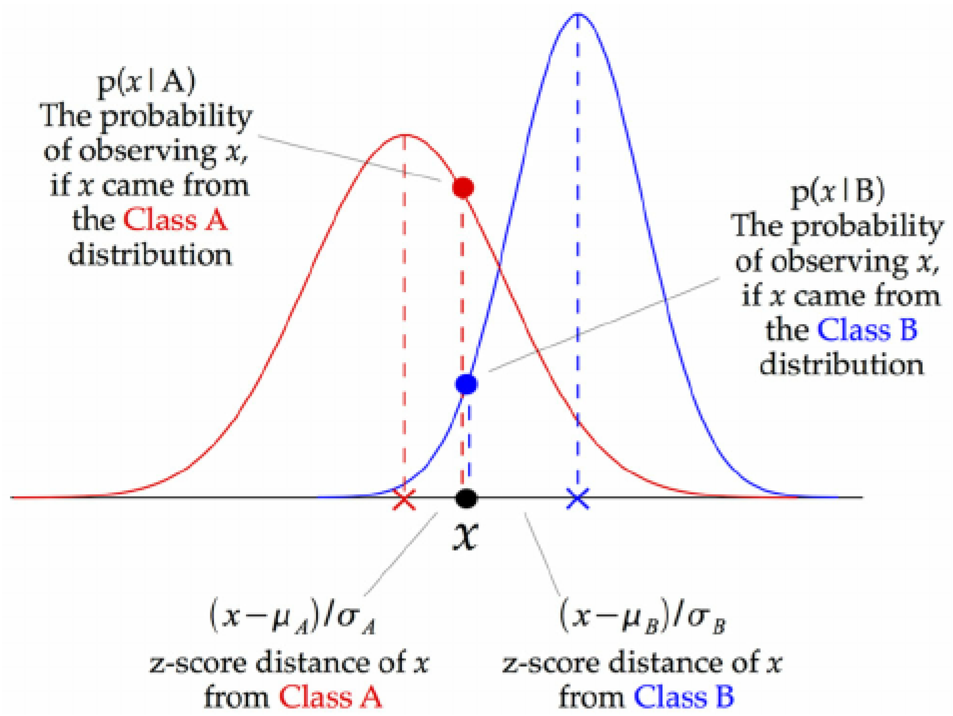
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

Class prior

Constant depending on Train data set size

Naive Bayes Model (1D)

Here Gaussian Bayes (other distributions possible!)



Bayes Rule:
Compute for all classes

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$$\boxed{f(x) := argmax_y P(x|y)P(y)}$$

Naive Bayes Classifier (ND)

$$P(y|x_1,...,x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

N-dim feature vectors

Important assumption:
Features are independent!

Here is the NAIVE assumption:

Naive Bayes Classifier (ND)

Final formulation:

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

Naive Bayes Classifier (ND)

Final formulation:

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

Discussion:

++ simple but powerful model
++ does not need much data
+ supports complex distributions (like mixture of Gaussians)
+ scales well

– – makes strong assumption on feature independence
– – estimation of complex distributions needs a lot of data

Discussion

Lab exercises coming up ...