Naïve Bayes

Probabilistic Classifier

Learn a function $f(x^{(i)})$ that uses a multidimensional input vector x to find a category $y_k^{(k)}$ among a set y of discrete and finite set of defined categories.

Where

$$x^{(i)} = \left[x_1^{(i)}, x_2^{(i)}, \dots \dots x_n^{(i)}\right]$$

and y defines a limited number of categories (e.g. win/lose/tie in a game)

Probabilistic classifier uses probability to pick the class with highest probability

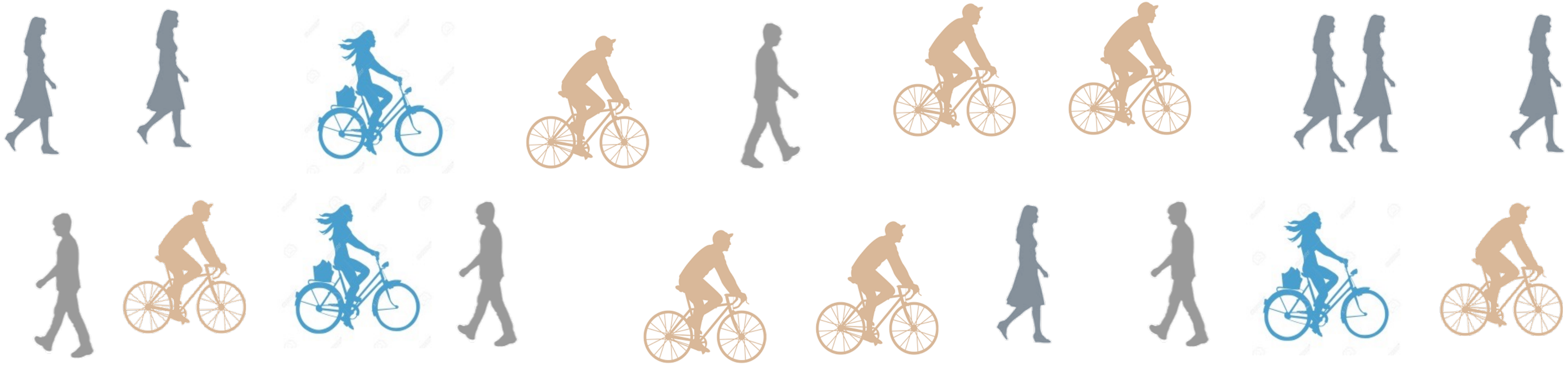
Bayesian Classifier

- Bayesian Classifier uses the Bayes rule to find the category y_k with the highest probability conditioned on x

Most probable category given observation

$$\hat{y} = \arg\max_{k}\left(\mathrm{P}(\mathrm{y_k}|\mathrm{x})\right)$$

Probability (review)



What is the probability that a person is a man ?

What is the probability that a person is a woman ?

Probability of an event

What is the probability that a person is riding a bicycle?

What is the probability that a person is riding a bicycle?

What is the probability that a woman is walking?

What is the probability that a man is riding a bicycle?

Joint probability of events

Dr.-Ing. Maggie Mashaly

Conditional Probability (example)

Only consider events that satisfy the condition in probability calculation

Given a person is a female

Given that a person is a female what is the probability that she is walking?

Given that a person is a female what is the probability that she is riding a bike?

$$\text{Probability of } (A|\text{female}) = \frac{Number\ of\ A\ and\ females}{\underbrace{Number\ of\ females}_{\text{Given}}}$$

Conditional probability (example)

Given a person is riding a bike

Given a person is riding a bike what is the probability that is person is a male ?

Given a person is riding a bike what is the probability that is person is a female?

Dr.-Ing. Maggie Mashaly

Example: Predicting the results of team based on weather

Temperature	Result
Hot	Win
Hot	Lose
Cold	Tie
Cold	Win
Hot	Win
Cold	Win
Hot	Win
Hot	Lose
Cold	Tie
Hot	Win
Hot	Win
Hot	Win
Cold	Tie

Temperature	Result
Cold	Win
Hot	Win
Cold	Tie
Cold	Win
Hot	Win
Cold	win
Hot	Win
Hot	Lose
Cold	Tie
Hot	Win
Cold	Lose
Cold	Tie
Cold	Tie

P(Win|Cold)=4/12

P(Lose|Cold)=1/12

P(Tie|Cold)=7/12

P(Win|Hot)=11/14

P(Lose|Hot)=3/14

P(Tie|Hot)=0/14

Conditional Probability (review)

- Probability of event B given the event A

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

Remember y is a single variable while x is a lager set of features

How do we calculate y given a large set of features?

- Probability of event A given the event B occurred

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

Easier to calculate but how to deal with $P(x, y)$

Bayes Rule (review)

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

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

$$P(y_k|x) = \frac{\overbrace{P(x|y_k)}^{\text{Class model}}\overbrace{P(y_k)}^{\text{Prior}}}{P(x)}$$

General Probability of the category (with out looking at the features

Bayes' Rule with multiple features

$$P(y|x_1, x_2, x_3 \ldots\ldots, x_n) = \frac{P(x_1, x_2, x_3 \ldots\ldots, x_n|y)P(y)}{P(x_1, x_2, x_3 \ldots\ldots, x_n)}$$

Assuming $x_1, x_2, \ldots. x_n$ are independent random variables

This the prediction of y given a set of feature

This is what makes the classifier naïve

Conditional independence
Assume y is the cause of the dependence

$$P(y|x_1, x_2, x_3 \ldots\ldots, x_n) = \frac{\prod_{j=1}^{n} P(x_j|y)\, P(y)}{\prod_{j=1}^{n} P(x_j)}$$

Discrete Naïve Bayes Classification example

- Consider an online shopping example where customers buy either books, movies or music. Customers of this site are described by gender, area, age category, education.
- For a new customer it is required to know if this customer would be interested in buying books, movies or music.
- Assume that input has n features $x = \{x_1, x_2, x_3 \dots \dots \dots, x_n)$, in this example:
  - $x = \{x_1 \rightarrow gender, x_2 \rightarrow area\ , x_3 \rightarrow age\ category\ , x_4 \rightarrow education\ )\}$
- Each feature is discrete and has k discrete possible outputs $x_j = \{x_{j,1}, x_{j,2}, \dots \dots x_{j,k}\ )$ where $x_{j,k}$ the outcome K for feature j in this example the gender has two possible outcomes (male and female) while feature age category has 4 possible outcomes (child, youth , Adult and old)
  - $x_1 = \{x_{1,1}(male), x_{1,2} \rightarrow (female\ )\}$
  - $x_3 = \{x_{3,1}(child), x_{3,2} \rightarrow (youth\ ), x_{3,3} \rightarrow (Adult), x_{3,4} \rightarrow (old)\ \}$
- The output y has 3 different possibilities (Books , Movies, Music)
- For each feature x_j calculate the conditional probability of feature options given the classification $P(x_j = q|y = c_1)$

Discrete Naïve Bayes Classification Algorithm Training Phase

For each possible classification output y_c

{

Calculate $P(y_c)$ the probability of y_c

Store the value of $P(y_c)$ versus y_c

For each feature of the input x_j

{

For each possible outcome k of feature j , $x_{j,k}$

{

Calculate $P(x_{j,k})$, the probability $x_{j,k}$.

Calculate $P(x_{j,k}|y_c)$ the probability $x_{j,k}$ given y_c.

Store $P(x_{j,k})$ and $P(x_{j,k}|y_c)$ versus $x_{j,k}$ in a table

}

}

}

Discrete Naïve Bayes Classification Algorithm Classification Phase

For a new customer with features x

For each possible classification y_c

{ Obtain $P(y_c)$ from the training tables

$P(y_c|x) = P(y_c)$

For each feature j of the input x_j

{ read the feature outcome k

Obtain the value of $P\left(x_{j,k}\right)$, the probability $x_{j,k}$ from the training tables

Obtain the value of $P\left(x_{j,k}|y_c\right)$ the probability $x_{j,k}$ given y_c from the training data $P(y_c|x) = P(y_c|x)\left(\frac{P\left(x_{j,k}|y_c\right)}{P\left(x_{j,k}\right)}\right)$

}

}

Compare all the values of $P(y_c|x)$ and pick the one with the highest probability to be the estimated classification

Training Phase Example

Training Data Set

Gender	Area	Age	Education	Classification
Female	A1	old	University	book
Female	A1	old	University	book
male	A1	old	High school	book
male	A1	Youth	High school	book
male	A1	Youth	High school	book
male	A1	Adult	University	movie
male	A1	Youth	University	movie
Female	A2	old	High school	book
Female	A2	Adult	University	book
male	A2	old	University	movie
Female	A3	old	University	book
Female	A3	Adult	University	movie
Female	A3	Adult	University	movie
male	A3	Adult	University	movie
Female	A4	Adult	High school	book
Female	A4	old	University	book
male	A4	Adult	High school	book
male	A4	Youth	University	book
Female	A4	Adult	University	movie
male	A4	Youth	University	movie

Class Probabilities

P(book)	0.6
P(movie)	0.4

Feature Probabilities And Conditional Probabilities

	P(feature outcome)	P(feature|book)	P(feature|movie)
P(male)	0.500	0.417	0.625
P(female)	0.500	0.583	0.583
P(A1)	0.350	0.417	0.250
P(A2)	0.150	0.167	0.125
P(A3)	0.200	0.250	0.375
P(A4)	0.300	0.333	0.250
P(Youth)	0.250	0.250	0.250
P(adult)	0.400	0.250	0.625
P(old)	0.350	0.042	0.125
P(high school)	0.300	0.500	0.000
P(university)	0.700	0.000	1.000

Classification Phase Example

P(book)	0.6
P(movie)	0.4

Find the classification for the following customer {male, A2, Adult, University}

$$P(book|customer) = \frac{0.417*0.167*0.25*0.5}{0.5*0.15*0.4*0.7} 0.6 \; = 0.414$$

$$P(movie|customer) = \frac{0.625*0.125*0.625*1}{0.5*0.15*0.4*0.7} 0.4 \; = 0.9301$$

	P(feature outcome)	P(feature|book)	P(feature|movie)
P(male)	0.500	0.417	0.625
P(female)	0.500	0.583	0.375
P(A1)	0.350	0.417	0.250
P(A2)	0.150	0.167	0.125
P(A3)	0.200	0.083	0.375
P(A4)	0.300	0.333	0.250
P(Youth)	0.250	0.250	0.250
P(adult)	0.400	0.250	0.625
P(old)	0.350	0.500	0.125
P(high school)	0.300	0.500	0.000
P(university)	0.700	0.500	1.000

Gaussian Bayesian Classifier (for continuous random variables)

$$P(y|x_1, x_2, x_3 \ldots\ldots, x_n) = \frac{\prod_{j=1}^{n} P(x_j|y)\, P(y)}{\prod_{j=1}^{n} P(x_j)}$$

How to calculate $P(x_j|y)$ if x_j *is continous*

Model each $(x_j|y)$ by a Gaussian distribution

Calculate the mean of every random variable x_j and the standard deviation of the random variable x_j and use a gaussian pdf with the same mean and standard deviation

Gaussian Bayesian Classifier

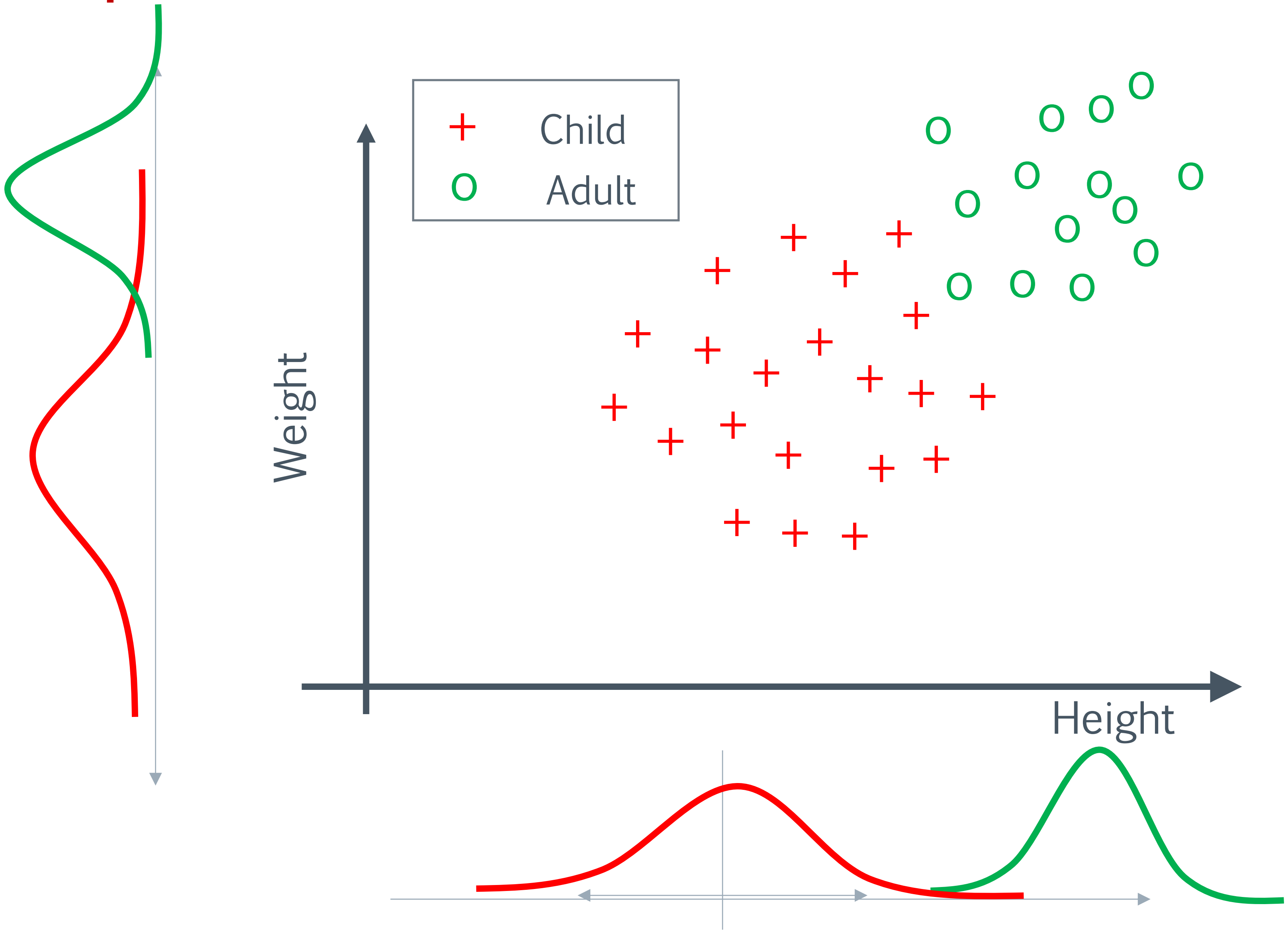
pdf of random variable representing $x_j|y$

$$P(y|x_1, x_2, x_3 \ldots\ldots, x_n) = \frac{\prod_{j=1}^{n} f_{(X_j|y)}(x_j|y)\,P(y)}{\prod_{j=1}^{n} f_{X_j}(x_j)}$$

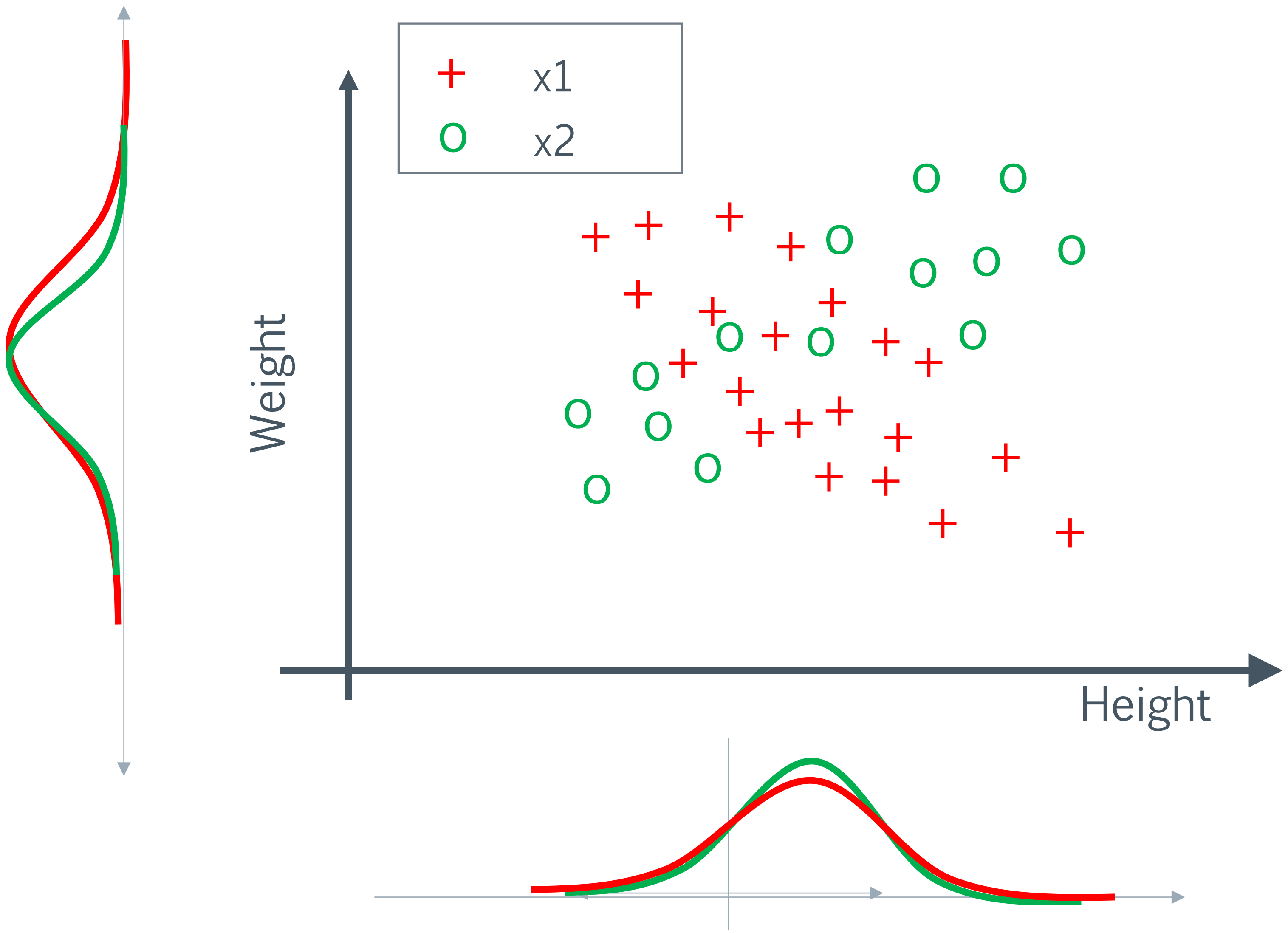
Pdf of random variable representing x_j

Ration of pdfs is equivalent to ratio of probabilities

Example



Failure of the naïve Bayesian classifier



Thank You