CLASSIFICATION I

Machine Learning for Autonomous Robots

Vamsi Krishna Origanti
Robotics Innovation Center, DFKI

November 9, 2023 – Bremen, Deutschland

University of Bremen

German Research Center for Artificial Intelligence

Questions

Any questions?

Recap

Recap questions

- What is Supervised Learning?
- What is classification?
- How does k-nearest-neighbours work?
- Explain Bayes' theorem. How can we use it for classification?
- What is logistic regression?
- How can we build decision trees efficiently?
- Why are Random forests useful?

Supervised Learning

Definition

Supervised learning is to learn by training an algorithm over example labeled data (input-output pairs), to be able to map an input space into an output space.

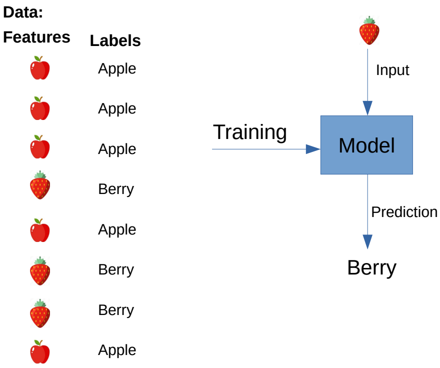

Figure: Supervised Learning Overview

Supervised Learning

- Consider the input variables x and the output variables Y
- Also, “n” training examples (x1, Y1), (x1, Y1), ... (xn, Yn)
- Aim is to find an appropriate mapping function “f” where Y = f(x).
- In simple words, the function f is a set of parameters q.
- The process of Learning q from the given data is nothing but “Training”. If output labels are already given for learning then its Supervised Learning.
- This training process yields a “model” which is a mapping function between inputs and outputs.
- If Y is a real value, the learning is called Regression.
- If Y is a categorical value, the learning is called Classification.

Naive Bayes example

Idea of classification

Intuitive Definition

We have some features of an instance and want to assign it a class.

$\Rightarrow$ We can only learn from some examples (life experience).

More formal definition

Supervised (e.g we have labels):

- Set of feature vectors $x \in X$
- Label y for each feature vector
- **We want:** Model that maps x onto y
- Using this model we can make then predictions for new/unseen data.

Classification in crime detection/prevention

Predict who might commit a crime.

Assume we have collected some features x_i for a set of known criminals (e.g. age, gender, religion, income, city ...).

For all criminals we also have the kind of crime they have committed (e.g. murder, fraud, traffic crime, sexual assault, robbery ...).

- Sun, Yao, Li, Lee *Detecting Crime Types Using Classification Algorithms* 2014: Journal of Digital Information Management.
- Sojaee, Mustapha, Sidi, Jabar *A Study on Classification Learning Algorithms to Predict Crime Status* 2013: International Journal of Digital Content Technology and its Applications.

Crime dataset as an example

- The latter paper uses the *Communities and Crime Data Set* from the UCI Machine Learning Repository, which consists partially from FBI UCR data of 1995.
- For our small example we use a **self-made artificial data set**.
- The first paper uses/compares the following classification algorithms: *Naive Bayesian, C4.5 (Decision Tree) and KNN.*

$\Rightarrow$ We are going to use **Naive Bayes** classifier for our example.

Using Bayes' theorem

Bayes' theorem

$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$

- Here: $P(A|B)$ is the likelihood of committed crime (A), given features (B).
- $P(B)$ is a constant so we can omit it.
- We also assume conditional independence (**naive**)
  $\Rightarrow P(B|A) = P(B_1|A) * \cdots * P(B_m|A)$
- Lets calculate the probabilities from our data (the more the better).

Our small crime dataset

Table: Small crime data set.

age	gender	income	city	comitted crime
27	Male	1200	Essen	No
35	Male	800	Munich	No
18	Female	0	Essen	Yes
15	Male	0	Bremen	Yes
25	Female	1700	Bremen	No
40	Male	2000	Essen	Yes
51	Male	1520	Ulm	No
26	Female	900	Munich	Yes

- This table is not representative and its intention is for the calculation of the example solely.

Our small crime dataset - quantised

Table: Small crime data set.

Age	Gender	Income	City	Comitted Crime
20-29	Male	1001-1500	Essen	No
30-39	Male	501-1000	Munich	No
10-19	Female	0-500	Essen	Yes
10-19	Male	0-500	Bremen	Yes
20-29	Female	1501-2000	Bremen	No
40-49	Male	1501-2000	Essen	Yes
50-59	Male	1501-2000	Ulm	No
20-29	Female	501-1000	Munich	Yes

- This table is not representative and its intention is for the calculation of the example solely.

Probabilities from our data set

age	crime	no crime
10-19	2/4	0/4
20-29	1/4	2/4
30-39	0/4	1/4
40-49	1/4	0/4
50-59	0/4	1/4

gender	crime	no crime
Male	2/4	3/4
Female	2/4	1/4

income	crime	no crime
0-500	2/4	0/4
501-1000	1/4	1/4
1001-1500	0/4	1/4
1501-2000	1/4	2/4

city	crime	no crime
Essen	2/4	1/4
Munich	1/4	1/4
Bremen	1/4	1/4
Ulm	0/4	1/4

- $P(\text{Crime}) = P(\text{no Crime}) = 4/8$

Testing a new instance

Testing a new instance

Okay let us test a new instance now. E.g., Xman[a] :)

[a] Similarities with existing persons is purely coincidental.

- Xman = (20-29, Male, 1501-2000, Bremen)
- Likelihood P(Crime|Xman) = P(C|Xman) ?
- Likelihood P(no Crime|Xman) = P(nC|Xman) ?

Let's use the naive assumption

- $P(C|Xman) = P(C) * P(20\text{-}29|C) * P(Male|C) * P(1501\text{-}2000|C) * P(Bremen|C)$
  $= 4/8 * 1/4 * 2/4 * 1/4 * 1/4 = 1/256 \approx 0.0039$
- $P(nC|Xman) = P(nC) * P(20\text{-}29|nC) * P(Male|nC) * P(1501\text{-}2000|nC) * P(Bremen|nC)$
  $= 4/8 * 2/4 * 3/4 * 2/4 * 1/4 = 3/128 \approx 0.023$

What are the results?

Is Xman a criminal?

So it is more likely that given his features, that he has not committed a crime than it is likely that he has committed a crime.

Or as a probability:

$P(C|Xman) = \frac{0.0039}{0.0039+0.023} \approx 14.5\%$

$P(nC|Xman) = \frac{0.023}{0.0039+0.023} \approx 85.5\%$

- As we can see, it is crucial to have enough data.
- Having a good model (from data) this can be used to predict if a certain person is likely to commit a crime and according measures in crime prevention can be undertaken.

Any further questions?