Foundations of Machine Learning - Homework 1
Group: L99

Justus Stahlhut, Salomo Pflugradt, Felix Lange

November 16, 2020

Exercise 1

a)

Supervised Learning:

Learn a function from a set of input-output-pairs.

Example: Learning to predict house prices.
To realize a function that maps houses to their respective prices, a dataset of (random) house-prices and their attributes is needed. An exemplary sufficient dataset could have information about the houses square footage, its number of rooms, features, and whether it has a garden or not. Given the dataset, a function *y(x)* can then be trained to predict house-prices for houses outside the dataset.

Unsupervised Learning:

Identify previously unknown structures/patterns in data with minimum human supervision.

Example: Determining customer segments in marketing data.
The goal might be to determine different groups of people that want to shop in an online store.
The main problem here lies in recognizing these types of people and their preferences (assuming they're given in the marketing data).
Given the dataset of different customers, applying clustering and labeling techniques splits the set into customer-groups that share important marketing-features. The categorization might then be used to propose articles a person might like based on articles other customers in the same cluster bought.

Reinforcement Learning:

Learn, adapt, or optimize a behavior strategy in order to maximize the own benefit by interpreting feedback that is provided by the environment.

Example: Autonomous lane changing.
Autonomous lane changing while driving can be realized by an algorithm called Q-Learning (Quality-Learning).
Given a so-called q-table of the systems current state and the following action (state, action) and q-values representing the most plausible action, Q-Learning fundamentally consists of these three steps:

1. Agent starts in state (s1), takes action (a1) and receives reward (r1)
2. Agent selects an action by referencing the Q-table with highest value OR acts randomly (with probability ϵ)
3. Updates q-values

Exercise 2

a)

(a1) *A pile of mushrooms - O*

(a2) *A table with the columns "size", "weight", and "color", as well as one row for each mushroom, and the respective measurements in the cells - X*

(a3) *A human mushroom expert, who can tell whether any mushroom you show them is poisonous or edible -* γ

(a4) *A device that measures size, weight and color of a mushroom -* α

(a5) *The set* $\{$*Poisonous, Edible*$\}$ *- C*

(a6) *The machine learning system that you are trying to build -* $c \approx y$

b)

(6a) *Annotating website classes -* γ

(6b) *Extracting features/information from HTML -* α

Exercise 3

a)

$$x = \begin{bmatrix} 5 \\ 7 \\ 15 \\ 28 \end{bmatrix}, y = \begin{bmatrix} 50 \\ 79 \\ 124 \\ 300 \end{bmatrix}$$

$$RSS(w_0, w_1) = \sum_{i=1}^{n} (y_i - w_0 - w_1 * x_i)^2$$
$$w_0 = \bar{y} - w_1 * \bar{x}$$
$$w_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\underline{w_0} = -7.48, \quad \underline{w_1} = 10.6$$

b)
$\hat{y}(Lada) = w_0 + w_1 * x_{Lada}$
$\hat{y}(Lada) = 151.52m$

c)

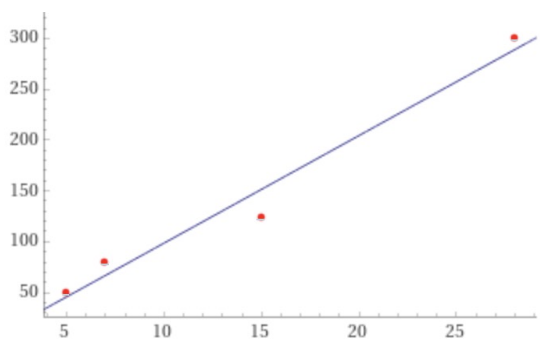
$$\mathbf{X} = \begin{bmatrix} 1 & 5 & 30\,530 \\ 1 & 7 & 90\,000 \\ 1 & 15 & 159\,899 \\ 1 & 28 & 270\,564 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 50 \\ 79 \\ 124 \\ 300 \end{bmatrix}$$

$$RSS(X) = (\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w})$$
$$\rightsquigarrow \mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\underline{\mathbf{w}} = \begin{bmatrix} -6.44 \\ 12.81 \\ -0.0002 \end{bmatrix}$$

$\hat{y}(Lada) = w_1 + w_2 * X_{Lada_{Age}} + w_3 * X_{Lada_{Mileage}}$
$\hat{y}(Lada) = 153,73m$

d)



e)
Drawbacks of extrapolation
One disadvantage is its unreliability when the given dataset has significant fluctuations. Extrapolation can still find the 'center' of these datapoints but the outcome might be too broad to use for for any research.

Another *major* drawback is that extrapolation assumes a given trend in the data will be followed - no matter what. While this type of behavior seems unrealistic in a real-world environment.

Exercise 4

a)
$A = \{A_1, ..., A_p\}$, A_i has m_i different values $\forall i \in \{1, ..., p\}$

$$n(p) = \prod_{i \in \{1,...,p\}} m_i$$

b)

$$|H_p| = \prod_{i \in \{1,...,p\}} (m_i + 1) + 2^{|A|}$$

c)

$$n(p+1) = m_p * n(p-1)$$

$$//H(p+1) = (m_p + 1) * H(p-1) + 2^{p-1}$$

Exercise 5

a)

Yes, it can. Given a large enough set of examples, there are prone to be people that enjoy sports when it's rainy and cold outside. Although the majority wouldn't enjoy getting wet, their examples aren't taken into account. Leading to an inconsistent hypothesis according to the data given by most people.

b)

Yes. *Regarding the example shown in the lecture:*
Following the order given by the examle: $x_1 \to x_2 \to x_3 \to x_4$, we can see that hypothesis h_4 is very general with 3 wildcard symbols: If we were to flip the example set, h_4 would be the most specific hypothesis, with h_3, h_2, h_1 being more general.