Lecture 1

Preprint · March 2019

1 author:

Alaa Tharwat
Fachhochschule Bielefeld

Learning from Data

Alaa Tharwat

Lecture 1: The learning problem

- Definition of machine learning
- Examples of machine learning
- Components of learning
- A simple linear model
- Types of learning

Lecture 1:

- Definition of machine learning
- Examples of machine learning
- Components of learning
- A simple linear model
- Types of learning

- Machine learning is a subset/part of the artificial intelligence science which uses statistical techniques to give computers the ability to *learn* to do a specific task without being explicitly programmed. "Arthur Samuel"
- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. "Tom Mitchell"

- Assume we have a set of pair numbers (2,4), (3,6), (4,8), and (5,10), and assume the first number in each pair is the input and the second one is the output
- We use the four pair numbers and put it into a machine to learn/train "how to predict" the second number (output)
- Can the trained machine learning model predict what is the output if the input is 6?. In other words, (6,**?**)
- In a handwriting recognition learning problem, what is E, T, and P?
  - E: data to collect which is collected
  - T: what decisions the software needs to make, i.e. recognizing and classifying handwritten words within images
  - P: how we will evaluate its results, i.e. percent of words correctly classified

Lecture 1:

- Definition of machine learning
- Examples of machine learning
- Components of learning
- A simple linear model
- Types of learning

- Assume we have a problem which is called *credit approval*.
- In this problem, given some features of some clients such as age, gender, annual salary, years in job, ..
- The goal of the model is to approve a new customer or not based on his/her data.

Customer	Age	Gender	Annual salary	Years in job	...
Thomas	32	M	50,000	2	...
Müller	45	M	60,000	9	...
Maren	58	F	35,000	15	...
..	...	...	...	...	..

- Given two types of fish, Salmon and Sea Bass
- In this problem, given some features about each type of fish such as width and weight
- The goal of the model is to estimate the type of fish given the width and weight value

Salmon

Sea Bass

Lecture 1:

- Definition of machine learning
- Examples of machine learning
- Components of learning
- A simple linear model
- Types of learning

- **Input**: ($\mathbf{x} \in \mathcal{R}^d$) (customer application, fish features), d is the number of features
- **Output**: ($y \in \{-1, +1\}$) (good/bad customer, type of fish)
- **Target function**: $f : X \rightarrow Y$ ((ideal credit approval formula, expert of fisheries), or the relation between X and Y. The target function (f) is unknown; therefore, we need to learn it
- **Data**: $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, where $y_i = f(x_i)$ (historical records, fish data)
- **Hypothesis**: $g : X \rightarrow Y$ (best hypothesis/formula to be used)

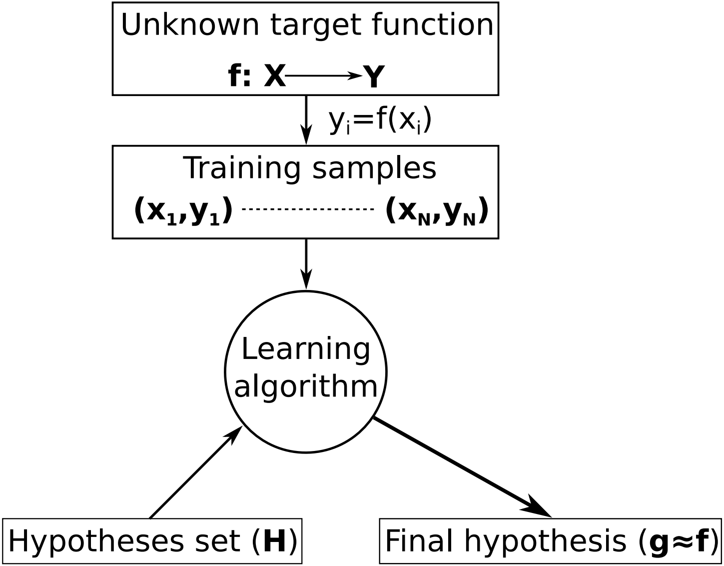
x, **y** and D are inputs to the learning problem, and f is fixed but unknown. The goal is to learn the function f from the data D.

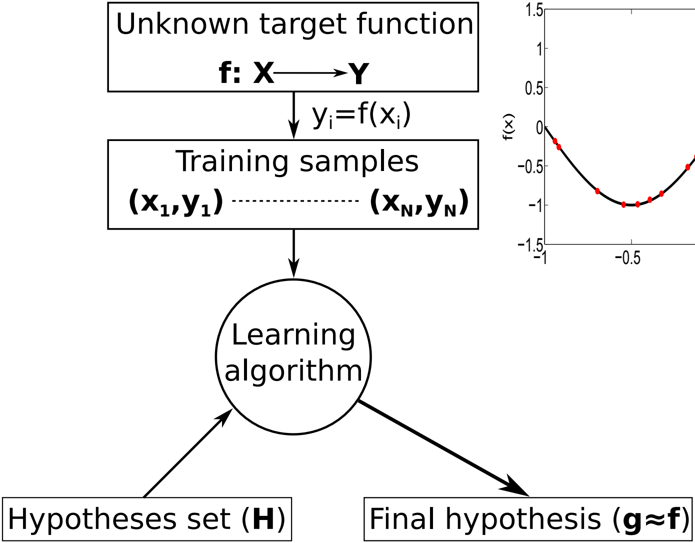
- In the fish example

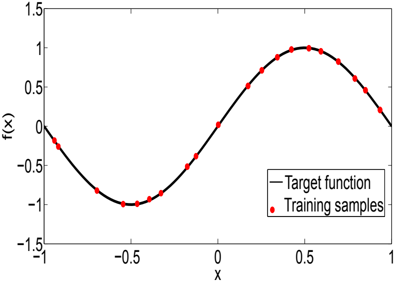
$$\mathbf{x} = \begin{bmatrix} 20 & 250 \\ 50 & 520 \\ 10 & 180 \\ 60 & 489 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} \tag{1}$$

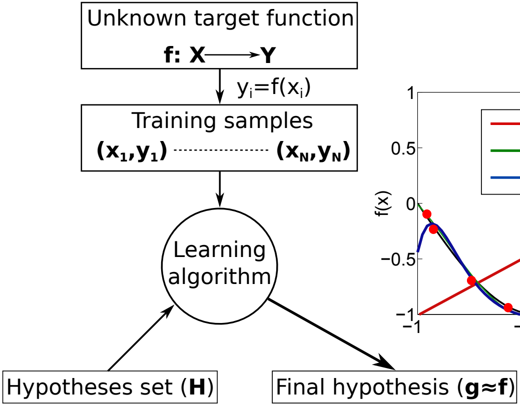
where

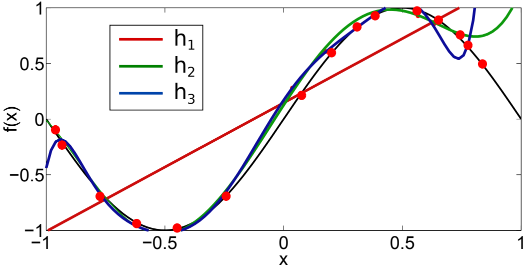
- the width is the first column in **x** and it is in mm
- the weight is the second column in **x** and it is in gm, i.e. in the first row/sample/object the fish is 20 mm width and 250 gm weight
- $y = 1$ means Salmon type and $y = 2$ means Sea Bass fish type


Unknown target function
f: X ⟶ Y
$y_i = f(x_i)$
Training samples
$(\mathbf{x_1},\mathbf{y_1})$ .... $(\mathbf{x_N},\mathbf{y_N})$
Learning algorithm
Hypotheses set (**H**)
Final hypothesis (**g≈f**)


Unknown target function
f: X → Y
yi=f(xi)
Training samples
(x1,y1) .... (xN,yN)
Learning algorithm
Hypotheses set (H)
Final hypothesis (g≈f)


f(x)
x
Target function
Training samples
1.5
1
0.5
0
−0.5
−1
−1.5
−1
−0.5
0
0.5
1


Unknown target function
f: X → Y
yi=f(xi)
Training samples
(x1,y1) (xN,yN)
Learning algorithm
Hypotheses set (H)
Final hypothesis (g≈f)


h1
h2
h3
f(x)
x
1
0.5
0
−0.5
−1
−1
−0.5
0
0.5
1

- The learning model consists of two components (solution components):
  - The learning algorithm such as Quadratic programming in Support Vector Machine (SVM)
  - The hypothesis set ($H = \{h_1, h_2, \dots, h_m\}$) (i.e. model). We need to select one hypothesis $g \in H$ that satisfies $g \approx f$.

Lecture 1:

- Definition of machine learning
- Examples of machine learning
- Components of learning
- A simple linear model
- Types of learning

- Given the input $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, where d is the number of features/attributes. It is also called, the dimension of the input. If $d = 1, 2$ or 3 we can visualize it, but, with high dimensions, the visualization is difficult
- In the perceptron learning algorithm, assume each feature has an importance factor or weight (w_i)

$$\begin{cases} \sum_{i=1}^{d} w_i x_i \geq \text{threshold} & \text{Approve credit} \\ \sum_{i=1}^{d} w_i x_i < \text{threshold} & \text{Deny credit} \end{cases} \tag{2}$$

- Using the above equation, the linear formula $h \in H$ can be written as follows:

$$h(x) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i x_i\right) - \text{threshold}\right) \qquad h \in H \quad (H \text{ is infinite}) \tag{3}$$

$$h(x) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i x_i\right) - \cancel{\text{threshold}}^{\,w_0}.\cancel{(+1)}^{\,x_0}\right)$$

$$h(x) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i x_i\right) - w_0\right) \tag{4}$$

$$h(x) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right) \tag{5}$$

- In vector form

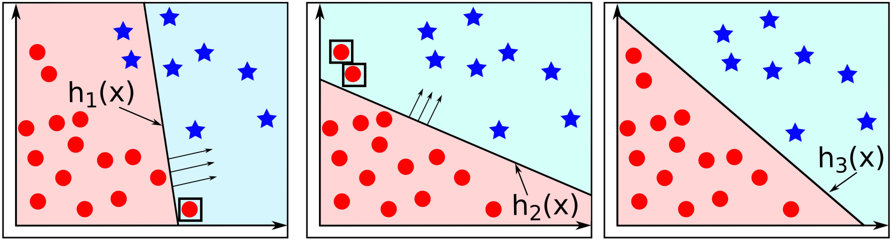
$$h(x) = \text{sign}(\mathbf{w}^T \mathbf{x}) \tag{6}$$

$$\mathbf{x} = \begin{bmatrix} x_0 = 1 \\ x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \in \mathcal{R}^{d+1} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} \in \mathcal{R}^{d+1}$$

- Again, w_0 is the bias.

How we can get many hypotheses?

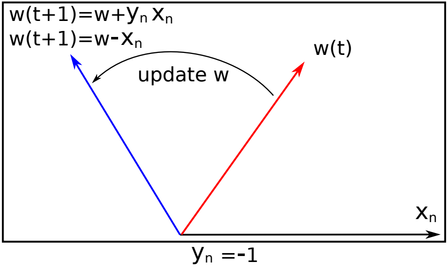
- By changing the model's parameter
  - e.g., changing **w** generates many hypotheses
- We want to select $g \in H$ so that $g \approx f$ on the data D

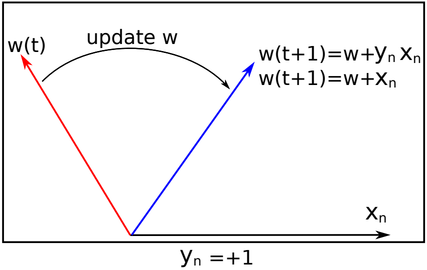


$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x}) \tag{7}$$

- pick a misclassified sample ((x_n, y_n)), this means that $\text{sign}(\mathbf{w}^T x_n) \neq y_n$
- update the weight vector as follows;

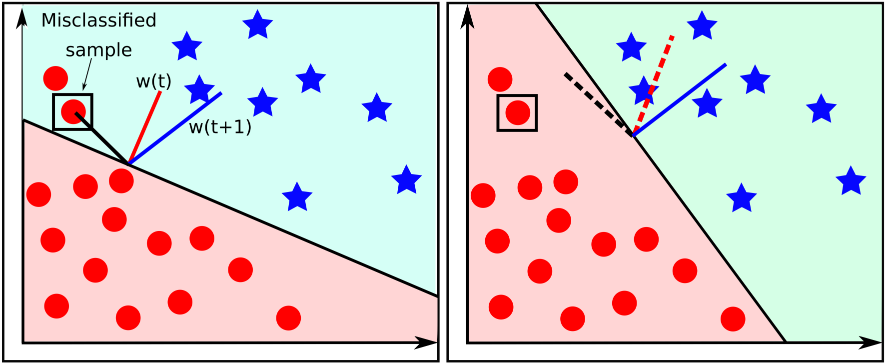
$$\mathbf{w}(t+1) \leftarrow \mathbf{w} + x_n y_n$$





How changing **w** changes the hypothesis?

- The figure below shows how one misclassified sample changes the weight (**w**) and hence the hypothsis move to make it correctly classified



Lecture 1:

- Definition of machine learning
- Examples of machine learning
- Components of learning
- A simple linear model
- Types of learning

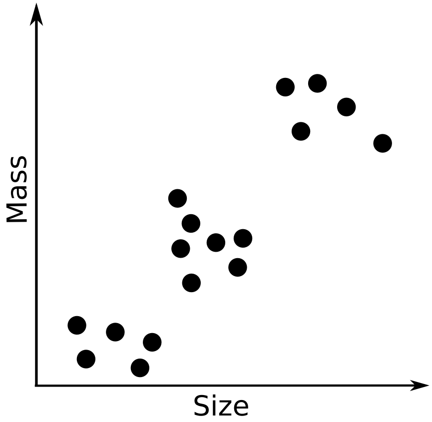
Machine learning tasks are classified into some categories based on whether there is a learning signal, teacher, expert, or feedback available to a learning system

- Supervised learning
- Unsupervised learning
- Reinforcement learning

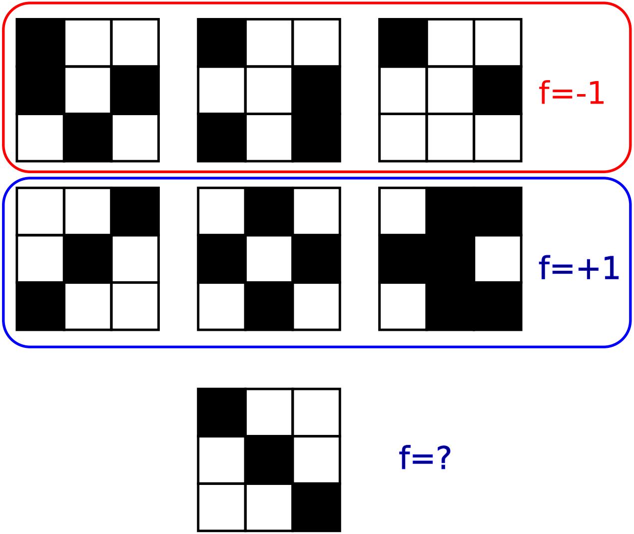
- In supervised learning, there is a data with our correct outputs
- These outputs/class labels guide the learning model to learn from the current data
- e.g. Assume we want to design a model for predicting the weather, and we have data. The data has some features/attributes (e.g. temperature, humidity, ...) and also class labels (e.g. sunny, cloudy, ...).
- There are two main categories in the supervised learning:
    - **Regression**: where the outputs are represented by real values
        - e.g. to predict the temperature in the weather prediction model
        - e.g. to predict the price of any item
        - e.g. to estimate the age of a person from his/her image
    - **Classification**: here the outputs are represented by class labels
        - e.g. to predict/identify a person (i.e. using face recognition or fingerprint system)
        - e.g. to predict whether a person is normal or patient

- In the unsupervised learning, there is no teacher or output to learn the model
- Clustering and Cocktail Party Algorithm are two main categories of the unsupervised learning
  - **Clustering**: The goal of clustering is to group the given data into groups/clusters. Each group has similar or related samples/objects
    - e.g. automatically group some documents into different clusters such as sports, politics, economics,...
  - **Cocktail Party Algorithm**: this allows you to find structure in a chaotic environment
    - e.g. identifying individual voices and music from a mesh of sounds at a cocktail party

Example of classifying coins



- In the Reinforcement learning, the goal of this type inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment to maximize some notion of cumulative reward.
  - Consider teaching a cat/dog a new trick. Normally, you cannot tell it what to do, but you can reward/punish it if it does the right/wrong thing. Similarly, we can train computers to do many tasks, such as playing backgammon or chess.
  - Teaching a game robot to perform better at a game by learning and adapting to the new situation of the game.
- So, instead of (input, correct output) in supervised learning, in reinforcement learning (input, some output, grade/reward for this output)


f=-1
f=+1
f=?