**Deep Learning**

# Section 1 – Meet your instructors and why machine learning

## 1.2.

1.2.1. What is the difference between artificial intelligence and machine learning?

a) They are the same

**b) Generally, machine learning is considered a part of artificial intelligence**

c) Generally, artificial intelligence is considered a part of machine learning

d) Artificial intelligence refers to hardware, while machine learning to software

Explanation: Artificial intelligence is intelligence shown by machines (opposed to natural (human) intelligence). Machine learning is a way to achieve artificial intelligence. AI is generally wider, and not so well-defined. There’s even a joke that AI is: “whatever hasn’t been done yet”.

1.2.2. Machine learning has been around for 60 years.

a) True

**b) False**

Explanation: While the term was coined in 1959 by Arthur Samuel, in fact, the first instances of “true” machine learning are much more recent. The breakthroughs in machine learning start in the ‘80s. In fact, the years prior to that are known as: ‘AI winter’, as there was very little funding going that way. It is important to realize that useful machine learning was not possible with the weak computers of the last century.

1.2.3. All translators are machine learning based.

a) True

**b) False**

Explanation: Google Translate relies on ML techniques, as well as traditional methods. Many other translators are simply dictionaries.

1.2.4. Which of the following is not an instance of machine learning?

a) voice recognition

**b) SPAM emails**

c) Netflix recommendation system

d) NSFW filters (programs that filter nudity, profanity, violence, etc.)

Explanation: While SPAM filtering is machine learning based for most big email providers, SPAM emails are sent by spammers. There are ways to automate these emails, but to the best of our knowledge SPAM emails are not machine learning based.

# Section 2 – Neural networks introduction

## 2.1. Introduction

2.1.1. Which of the following is NOT a building block of a machine learning algorithm?

a) Data

**b) Variable**

c) Objective function

d) Optimization algorithm

Explanation: The four ingredients are data, model, objective function, and optimization algorithm. While there are variables, they are a part of the model.

## 2.2. Training the model

2.2.1. Training the model is:

a) A pure trial and error process

**b) A kind of trial-and-error process with some feedback**

c) The process of giving guidelines to the computer how to find patterns

d) The process of watching people or machine perform an activity and replicating them

Explanation: The training process is essentially a trial-and-error process, but each consequent trial is better than the previous one, as we have methods in place that give feedback to the algorithm, as we will explore later.

2.2.2. A machine is easily better at making coffee than a human.

a) True

**b) False**

Explanation: Each machine learning algorithm that has great successes like making very good coffee is quite complicated. It is easier said than done.

2.2.3. Self-driving cars learn by:

a) Driving for as many hours as possible before learning how to do it safely and efficiently

b) A very strict set of rules that Elon Musk and the others are programming day and night

**c) “Watching” thousands of hours of footage of real people driving**

d) Breaking the rules (e.g. go on red light, go over the speed limit) and waiting to get punished for it

## 2.3. Types of supervised learning

2.3.1. What is NOT a type of machine learning?

a) supervised

b) unsupervised

**c) reinforced**

d) reinforcement

Explanation: The three types of machine learning are called: supervised, unsupervised, and reinforcement. While some sources don’t recognize reinforcement as a distinct type, it is “different enough” to separate it.

2.3.2. You have 10,000 photos of cars and 10,000 photos of ships. You label each of them as either a car or a ship, and create a machine learning algorithm that, based on the labels you assigned, must classify them. What type of machine learning is that?

**a) supervised**

b) unsupervised

c) reinforcement

Explanation: As the photos are labeled, we are talking about supervised learning.

2.3.3. Recently, Google created a 3D model which was running around obstacles. It had to find the best path, so that it reaches the end of the track taking as little “damage” as possible. The model tried going through the track on 2 legs so walking, on 2 legs and 2 hands, so crawling, and so on. It had a chance to fall off or be stopped at an obstacle. At the end it would be assigned a score, which reflected how well it did. This is an instance of:



a) supervised learning

b) unsupervised learning

**c) reinforcement learning**

Explanation: Similar to the Super Mario example, the algorithm learns through the positive feedback it receives. The higher the score, the better it did.

2.3.4. You have financial data for 100 countries. You feed them to the algorithm and ask it to classify them in as many groups as it sees fit. It starts with 100 groups as each country represents a separate group. You decide to tell it to spit out 5 major groups, i.e. cluster them in 5 clusters. This is an instance of:

a) supervised learning

**b) unsupervised learning**

c) reinforcement learning

Explanation: Although you told the algorithm you need 5 clusters, you did not help it in any way. It classified them based on financial similarities, rather than labels. Clustering is almost always unsupervised.

## 2.4. The linear model

2.4.1. The linear model is given by:

a) y = xTw + b

b) y = wx + b

c) y = wT x + b

**d) All of the above**

Explanation: It doesn’t matter how you define the linear model. Different fields use different notations. In this course, will actually use the notation y = xw + b.

2.4.2. Using the linear model, and knowing that the weight is 0, the bias is -5. Given an x = 10, what’s the value of y?

a) 10

**b) -5**

c) 5

d) -50

Explanation: Notice that when the weight of a linear model is 0, we don’t care about the input information we have.

## 2.5. The linear model. Multiple inputs

2.5.1. The linear model for multiple inputs is given by:

**a) y = xw + b**

b) y = x1w1 + x­2w2 + b

c) y = x1w2 + x2w1 + b

d) y = x2w2 + x2w2 + b

Explanation: No matter the number of inputs, the model doesn’t change. The difference is that x and w are vectors, rather than scalars. Notice that d) is plainly wrong, while b) and c) are examples of linear models with only 2 inputs. In general, we can have infinitely many inputs.

2.5.2. You have y = xw + b, where w = [1.2, -3], while b = [7]. If x = [2 , 3], what is the value of y?

**a) 0.4**

b) – 13.6

c) 4.6

d) 9.4

## 2.6. The linear model. Multiple inputs and outputs

2.6.1. The linear model for multiple inputs and outputs is given by:

a) ny = nxw + nb

**b) y = xw + b**

c) [y1, y2] = xw + b

Explanation: Same as the quiz question for the previous lecture. Although these questions may seem stupid, we must emphasize that y = xw + b is all you need.

2.6.2. If the number of inputs is k, the number of outputs is m, and the number of observations n, what is the shape of the weights matrix, w?

a) n x m

b) n x k

**c) k x m**

d) m x k

Explanation: The size of the weights is the number of inputs times the number of outputs, so k x m. It is extremely important to understand that the weights don’t depend on the number of observations (samples).

2.6.3. If the number of inputs is k, the number of outputs is m, and the number of observations n, what is the shape of the biases, b?

a) n x m

**b) 1 x m**

c) 1 x k

d) n x k

Explanation: The number of biases depends on the number of output variables. Moreover, like the weights, the biases don’t depend on the number of observations.

2.6.4. How are the parameters called in machine learning?

a) coefficients and intercepts

b) coefficients and biases

c) weights and intercepts

**d) weights and biases**

## 2.7. A picture

2.7.1. Data which can be classified using a linear model is called:

a) clusterable

b) regressable

c) linear

**d) linearly separable**

Explanation: There is no word clusterable. Same for regressable, although a bit controversial. “Linear data” has no meaning, at least to us.

## 2.8. The objective function

2.8.1. In supervised learning, we are dealing with:

a) lost functions

**b) loss functions**

c) reward functions

d) reinforcement functions

Explanation: In supervised learning, we are usually measuring the error of prediction, and we use loss functions.

2.8.2. Reward functions are NOT:

a) functions we are trying to maximize

b) functions used in reinforcement learning

**c) functions we are trying to minimize**

d) functions

Explanation: Functions we are trying to minimize are called loss functions.

## 2.9. L2-norm loss

2.9.1. A target is:

**a) The correct value at which we are aiming**

b) A synonym for output

c) A part of the model

d) Always bigger than 0

Explanation: The targets are the labels, so we are 100% sure they are the correct values, which our model should learn to predict. We use them to train the model, but they are not a part of the model itself, as a linear model is defined solely by its parameters. A target may be negative.

2.9.2. The objective function measures:

a) how well the targets match our model’s outputs

**b) how well our model’s outputs match the targets**

c) the model’s parameters

d) linearity of the data

Explanation: The targets are the labels, so they are always the same. We are trying to obtain outputs, which are closest to the targets, so it is more correct to say that the objective function measures how well the outputs match the targets.

2.9.3. The L2-norm loss is used for:

a) k-means clustering

b) classification

**c) regression**

d) hierarchical clustering

## 2.10. Cross-entropy loss

2.10.1. Cross-entropy loss is used for:

a) k-means clustering

**b) classification**

c) regression

d) hierarchical clustering

2.10.2. Which cross-entropy points to the best match between outputs and targets?

a) L = 12.41

b) L = 0.78

**c) L = 0.44**

d) L = 0.77

Explanation: The lower the loss function, the higher the accuracy of prediction.

2.10.3. The cross-entropy loss divided by 10 is not a loss function that can be used for machine learning

a) True

**b) False**

Explanation: Any function that holds the basic property of being higher for worse results, and lower for better results can be a loss function. Dividing by some constant changes nothing.

2.10.4. The cross-entropy loss MULTIPLIED by 10 is not a loss function that can be used for machine learning

a) True

**b) False**

Explanation: Any function that holds the basic property of being higher for worse results, and lower for better results can be a loss function. Multiplying by some constant changes nothing.

## 2.11. 1-parameter gradient descent

2.11.1. The gradient is:

**a) a generalization of the derivative concept**

b) a generalization of the integral concept

c) a generalization of the optimization algorithm

d) a generalization of the objective function

Explanation: The gradient is a mathematical term. It is the multivariate generalization of the derivative concept.

2.11.2. The gradient descent is a type of:

a) data

b) model

c) objective function

**d) optimization algorithm**

Explanation: The gradient descent is not the only way to optimize an algorithm. However, it is the most basic.

2.11.3. The learning rate is denoted by which Greek letter?

a) alpha

b) sigma

c) nabla

**d) eta**

2.11.4. A high learning rate:

a) is better, as you learn faster

**b) is faster, but may not reach the minimum**

c) is slower, but sure to reach the minimum

d) is faster, and always reaches the minimum

Explanation: A too high learning rate may cause the loss function to diverge to infinity, instead of finding the minimum.

## 2.11. N-parameter gradient descent

2.12.1. N-parameter gradient descent differs from the 1-parameter gradient descent as it deals with:

**a) many weights and biases**

b) many input variables

c) many output variables

d) many targets

Explanation: The N-parameter GD updates many weights and biases. The 1-parameter GD still could have many inputs, outputs and targets, but related to a single weight.

2.12.2. We use the delta to denote:

a) difference in models

b) difference between outputs and inputs

**c) difference between outputs and targets**

d) difference between methodologies

2.12.3. The weights and biases:

a) Are the same thing

b) Have completely different update rules

c) Are rarely updated

d) **Have update rules following the same logic**

Explanation: The update rules may differ a bit, but follow the same logic of updates (derived from the 1-parameter gradient descent lecture).

# Section 3 – Setting up the environment

## 3.2. Why Python and why Jupyter

3.2.1. What is “Jupyter”?

a) A software that transforms Python and some other languages into 1s and 0s

**b) A server-client application**

c) A specific group of language kernels

d) An environment where language kernels are related to specific hardware parts

3.2.2. What is the default type of file in Jupyter that allows us to store Python code?

a) Regular Notebook type of file

b) Dot py(\*.py)

**c) IPython Notebook file (\*.ipynb)**

d) None of the above

## 3.5. Jupyter dashboard

3.5.1. Which of the following options cannot be applied to an IPython Notebook file that is currently running?

**a) Rename it**

b) Shut it down

c) Duplicate it

d) Delete it

3.5.2. Which type of cell does not exist in Jupyter?

a) Code

b) Markdown

**c) Picture**

d) None of the above

3.5.3. Which is the shortcut combination for deleting a selected cell in a Notebook Document?

a) “A”

b) “B”

**c) “D” Twice**

d) “X”

## 3.6. Installing packages

Problem: Install two packages, if you don’t have them.

1. Sklearn – a machine learning package
2. Pandas – a powerful data manipulation and analysis package, which you may need out of the scope of this course

Solution:

1. Enter the Anaconda cmd, as seen in the lecture and write:

pip install sklearn

or

conda install sklearn

1. Enter the anaconda cmd, as seen in the lecture and write:

pip install pandas

or

conda install pandas