Task 1

ACM Summer School 2021, Artificial Intelligence

Mentors: Mobina Shahbandeh, Amirhossein Abaskohi,
Mahdi Mehmanchi, Erfan Mirzaei

1 Introduction

In this task we will extract information from a CSV file. This file (*diabetes_dataset.csv*) includes medical information of patients who have/do not have Diabetes.

2 Steps

All of the steps should be done using **vectorization**. Pandas includes a generous collection of vectorized functions for everything from mathematical operations to aggregations and string functions (for an extensive list of available functions, check out the Pandas docs). You can read more about vectorization in this link.

2.1 Read the file and save it as a DataFrame.

2.2 Show the first and last *five* rows of the DataFrame.

2.3 Show mean of *age* and standard deviation of *blood pressure*.

2.4 There are some empty cells (called missing data) in the dataset. Fill them with mean of the corresponding column.

2.5 Show how many patients have and how many do not have Diabetes.

2.6 Sort the patients based on their *age* in ascending order.

2.7 Select the patients who are older than *60* (inclusive).

2.8 Show mean of each column based on *outcome* (having/not having Diabetes).

2.9 Draw the histogram of the columns in subplots.

2.10 Plot the correlation of the columns.