Alexandria University
Faculty of Engineering
Computer and Systems Engineering
Department

CS435 Introduction to Deep Learning
Assigned: Monday, February $17^{th}$, 2019
Due: Wednesday, March $6^{th}$, 2019

**Assignment 1**
**Linear Regression**
**&**
**Logistic Regression**

# 1 Objective

- Getting to know Keras platform.

- Getting to know how to use Google Colaboratory.

- Building a linear regression model.

- Building a logistic regression model.

# 2 Introduction to Keras

- Keras is a high-level neural network API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano.

- We will be concentrating on using Keras on top of TensorFlow

- All the information needed to use Keras API can be found in this documentation https://keras.io/

# 3 Introduction to Colaboratory

- Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

- You can get all the information you need about how to use Colaboratory from this link https://colab.research.google.com/notebooks/welcome.ipynbscrollTo=-Rh3-Vt9Nev9

# 4 Linear Regression

The problem we are trying to solve here is finding a new house which is suitable to our needs and the budget we assigned. The client who wants to buy the new house did her research and found some houses. She wrote the details of each house she visited including location, sale condition, sale type, house price, among others. She needs some help to know how much she is expected to pay to get a house that conforms with her specific needs.
Your task is to build a linear regression model that helps her to predict the house price depending on the given attributes she collected.

Alexandria University
Faculty of Engineering
Computer and Systems Engineering
Department

CS435 Introduction to Deep Learning
Assigned: Monday, February $17^{th}$, 2019
Due: Wednesday, March $6^{th}$, 2019

## 4.1 Dataset

- You will find the training and testing dataset files (housing_train.csv and housing_test.csv) in the resources section on Piazza.

- You will find the starter code in assignment1_housing.py.

- This dataset contains 80 features that demonstrate the state of the house and our target which is the house price.

# 5 Logistic Regression

Here, we are trying to increase the peoples' attention regarding the heart diseases. Like any disease, it is always better to know if you are sick early so you can get the treatment you need before it is too late. Therefore, we use a dataset that gathered some information about two groups: a group with a heart disease and the other group has no disease.
The gathered information includes age, chest pain type, fasting blood sugar, etc. Your goal is to train a logistic regression model to predict if a person has a heart disease or not depending on the given information.

## 5.1 Dataset

- You will find the dataset file (*heart.csv*) in the resources section on Piazza.

- You will find the starter code in assignment1_heart.py.

- This data set contains 13 features that demonstrate the health state of a person and our target (0 if this person does not have a heart disease and 1 if he has a heart disease.)

# 6 Requirements

- You should build the model for each problem. You are expected to complete the missing part of the code and expected to understand the **whole** code.

- You should try to change the model's parameters in order to get better performance on the test set.

    - List the hyper-parameters of your model and their effect on performance.
    - Demonstrate the effect of adding artificial features to your model.
    - Check your model performance using different loss functions.
    - Demonstrate the effect of using different optimizers and which one produced better performance.

Alexandria University
Faculty of Engineering
Computer and Systems Engineering
Department

CS435 Introduction to Deep Learning
Assigned: Monday, February $17^{th}$, 2019
Due: Wednesday, March $6^{th}$, 2019

- – Show if your model is over-fitting or under-fitting and explain how can you solve these problems.

- You should plot the training, validation and testing accuracies to better understand the performance of your model.

# 7 Bonus

Your models will be tested against a private dataset. The highest three performing models will take the bonus grade.

# 8 Notes

- Parts of this assignment are based on Kaggle kernels.

- The dataset files and the starter code for the problems can be found in the resources section in Piazza.

- You should deliver a report explaining all your work.

- Cheating will be severely penalized (for both parties). So, it is better to deliver nothing than deliver a copy. Any online resources used must be clearly idenified.