# Introduction to Machine Learning (Spring 2019)

# Homework #1 (Due date: April 8)

**Student ID** _____

**Name** _____

**Instruction:** We provide all codes and datasets in Python. Please write your code to complete two models: linear regression and logistic regression. Besides, please measure the performance for each model.

**(1) [30 pts]** Implementation

(a) **[Linear reression]** Implement training and evaluation function in 'models/LinearRegression.py' ('train' and 'eval' respectively).

**Answer: Fill your code here. You also have to submit your code to i-campus.**

(b) **[Logistic reression]** Implement training and evaluation function in 'models/LogisticRegression.py' ('train' and 'eval' respectively).

**Answer: Fill your code here. You also have to submit your code to i-campus.**

(c) **[Optimization]** Implement SGD, Momentum, RMS Prop optimizers in 'optim/Optmizer.py'. Training should be based on the minibatch, not the whole data.

**Answer: Fill your code here. You also have to submit your code to i-campus.**

NOTE: You should write your codes in 'EDIT HERE' signs. It is not recommended to edit other parts. Once you complete your implementation, run the main codes to check if it is done correctly ('linear_main.py' for Linear Regression and 'logistic_main.py' for Logistic Regression).

**(2) [30 pts]** Experimental results

(a) **[Linear Regression]** For 'Graduate' and 'Concrete' dataset, adjust the number of training epochs and learning rate to minimize RMSE. Report your best results for each optimizer.
(Batch size = 10, epsilon = 0.01, gamma = 0.9)

**Answer: Fill the blank in the table.**

| Dataset | Optimizer | # of epochs | Learning rate | MSE |
|---|---|---|---|---|
| **Graduate** | SGD | | | |
| | Momentum | | | |
| | RMSProp | | | |
| **Concrete** | SGD | | | |
| | Momentum | | | |
| | RMSProp | | | |

(b) **[Logistic Regression]** For 'Titanic' and 'Digit' dataset, adjust the number of training epochs and learning rate to maximize accuracy. Report your best results for each optimizer.
(Batch size = 10, epsilon = 0.01, gamma = 0.9)

**Answer: Fill the blank in the table.**

| Dataset | Optimizer | # of epochs | Learning rate | Acc. |
|---|---|---|---|---|
| **Titanic** | SGD | | | |
| | Momentum | | | |
| | RMSprop | | | |
| **Digit** | SGD | | | |
| | Momentum | | | |
| | RMSprop | | | |

(c) **[Optimization]** For 'Titanic' dataset, execute the logistic regression with three optimization methods. Given the following parameter settings, draw two plots : a plot whose x-axis and y-axis are epochs and accuracy, and a plot whose x-axis and y-axis are epochs and cross-entropy loss. Explain which optimization method shows the best accuracy.

| **Parameter Settings** | |
|---|---|
| Batch size | 10 |
| Learning rate | 0.0005 |
| Epsilon | 0.01 |
| Gamma | 0.9 |
| # of Epochs | 30, 60, 90, …, 300 |

**Answer: draw the plot and explain the result, especially about the correlation with loss and accuracy according to different optimization methods.**