

## Assignment 5 - OVERFITTING

Submit file PDF: [assign5\\_lớp.nhóm\\_tênho.pdf](#)

For example: [assign5\\_02clc.03\\_quetd.pdf](#)

[Assign5\\_02cq.03\\_quetd.pdf](#)

- Submit before **MONDAY (22/09) 11:30PM for CHINH QUY**
  - Submit before **WEDNESDAY (24/09) 11:30PM for CLC classes**
- =====

### 5.1. (5 pages) Overfitting

- What is overfitting? Copy image from textbook\_2 and explain
- Techniques for decreasing/preventing overfitting: *present each technique in detail*
- Examples with codes

5.2. Using 2 techniques *change learning rate, augmentation* to decrease overfitting of models CNN, RNN, LSTM given in **YOUR Assignment 4.2**. Analyze the results given by output

### 5.3. (5 pages) Presenting 10 techniques for representing text in vector/tensor

- Onehot (Word counting)
- bagOfWord
- Tf-idf
- Word2vector
- Bert
- .....

### 5.4 Text—tensor/vector

- a. Load dataset of film review  
200 real (well-known) film titles.  
500 users (user\_001 ... user\_500).  
5,000 reviews (one per row) with fields:  
review\_id, user\_id, user\_name, film\_title, rating (1-10), review\_text,  
review\_date.  
given on facebook site filmReview\_1.csv
- b. Using BERT and word2vector to represent vector/tensor
- c. Using CNN, RNN, LSTM to analyze review dataset to determine trends of film, preference film....
- d. Your models are overfitting or not. If overfitting, using techniques change learning rate and augmentation to improve

=====

## ASSIGNMENT 4.2

### 4.2. Extend Assignment 4.1

- a. Build a dataset with **10.000 persons and 20 features** (jobs, age, gender, height, weight....) and store in C:\DATA\data\_4.2.csv
- b. Show the distribution of the dataset
- c. Using 5 basic ML models and using BMI to classifying and predicting

- d. Build models CNN, RNN, LSTM with  $\geq 5$  layers for classification and prediction problem in overweight, underweight, normal with respect to *age, job, gender, area, diet*
- e. Compare and evaluate models given with metrics accuracy, MAE, MSE, RMSE
- f. Visualize the results:
  - Which age has more overweight persons?
  - Which job has more overweight persons?
  - Which gender has more overweight persons?
  - Which area has more overweight persons?
- g. Deploy the best model so that user enters features and output is underweight, overweight or normal
- h. Build a BIG Knowledge base more 3 M and store in C:\DATA\kb\_healthGuide.json and give a guide for health based on this knowledge base.
- i. Build guide with chat based on your knowledge base and prediction result