# Knowledge Mining and Big Data Part 2

Pedram Ghazi

Student Number: 267640

1) A) One major difference among incremental learning and batch learning techniques is that in incremental learning classifier is trained(updated) after each incoming instance while in batch learning data-set is used all at once.

   B) We use each instance first to test the model, and then to train the model. From this the accuracy can be incrementally updated. This method takes the advantage of the fact it does not need extra data-set for testing so it makes maximum use of the available data.

2) A)

| | | |
|---|---|---|
| Correctly Classified Instances | 74748 | 74.748 % |
| Incorrectly Classified Instances | 25252 | 25.252 % |

B)

| | | |
|---|---|---|
| Correctly Classified Instances | 30662 | 16.9392 % |
| Incorrectly Classified Instances | 150350 | 83.0608 % |

C)

| | | |
|---|---|---|
| Correctly Classified Instances | 68502 | 68.502 % |
| Incorrectly Classified Instances | 31498 | 31.498 % |

D)

| | | |
|---|---|---|
| Correctly Classified Instances | 36701 | 20.2755 % |
| Incorrectly Classified Instances | 144311 | 79.7245 % |

E)

| | | |
|---|---|---|
| Correctly Classified Instances | 54099 | 54.099 % |
| Incorrectly Classified Instances | 45901 | 45.901 % |

F)

| | | |
|---|---|---|
| Correctly Classified Instances | 65941 | 36.4291 % |
| Incorrectly Classified Instances | 115071 | 63.5709 % |

G)

| | | |
|---|---|---|
| Correctly Classified Instances | 50297 | 50.297 % |
| Incorrectly Classified Instances | 49703 | 49.703 % |

H)

| | | |
|---|---|---|
| Correctly Classified Instances | 88685 | 48.994 % |
| Incorrectly Classified Instances | 92327 | 51.006 % |

3) As the amount of training instances increases, performance becomes better:
   In question 2-b) 16.9392 %
   In question 2-d) 20.2755 %
   In question 2-f) 36.4291%
   In question 2-h) 48.994 %

4) This part was an example for incremental training and testing of a naive Bayes classifier and in this we understood whenever we have an incoming dataset, performance becomes better and more accurate for the same test dataset.