BIA-660 **Kai Zhang**

Q1 comments:

- We have known that the best f1 score from grid search is 0.8753..., however, f1-macro score obtained from the test dataset is 0.90 which is greater than previous one. Therefore, the model is overfitted!

Q2 classifier performance analysis

1) How sample size affects each classifier's performance?

    1. Generally, two classifiers' performance becomes better as sample size is increasing although there are some ups and downs.

2) How many samples do you think would be needed for each model for good performance?

    1. At least 6000 samples since when sample size is greater than 6000, not only Bayes model but also SVM model can reach 90% or more on precision and recall.
    2. For Bayes model, it can reach good performance although there are only about 3000 samples.

3) How is performance of SVM classifier compared with Naïve Bayes classifier, as the sample size increases?

    1. All in all, the Bayes model's performance is always better than SVM model.
    2. When sample size is less than 1800, the performance of two models are similar. They are all increasing steadily.
    3. When sample size lies in interval between 1800 and 4200, Bayes model performs much better than SVM model. And then, their performance is very close.