Report
INF8245E (Fall 2021) : Machine Learning - Assignment #3

Amine EL AMERI - Matricule: 2164634

Medical Text Classification

1- I converted the dataset to binary bag-of-words representation and frequency bag-of-words representation and submitted the required datasets.

2- Using Medical-NLP dataset with BBoW representation and evaluating with the F1-score

a) After I implemented the random classifier and the majority-class classifier
The F1 score of the random classifier is 0.261
The F1 score of the majority classifier is 0.121

b) I implemented Bernoulli Naive Bayes, Decision Trees, Logistic regression and Linear SVM

c) The list of hyperparameters for each classifier:

Classifier	Hyperparameters and their range	Best values of hyperparameters
Bernoulli Naïve Bayes	Smoothing parameters alpha Tested the range [0, 1] with a step size 0.05	alpha = 0.65
Decision Trees	The criterion, tested gini and entropy The max depth of the tree, tested {10, 30, 50, 70} The minimum number of samples required to split an internal node, tested {0.1, 0.3, 0.5, 0.7, 0.9}	criterion: gini max_depth: 30 min_samples_split: 0.1
Logistic Regression	The norm of the penalty, tested l1 and l2 The regularization parameter, tested {0.01, 0.05, 0.1, 1, 5, 10, 50}	Penalty: l1 Regularization param C = 0.1
Linear SVM	The regularization parameter, tested the range [0.001, 0.1] with a step size 0.001	Regularization param C = 0.093

d) The training, validation, and test F1-score for all the classifiers

Classifier	F1-score with the best parameters
Bernoulli Naïve Bayes	BernoulliNB train f1_score with best param: 0.537 BernoulliNB valid f1_score with best param: 0.460 BernoulliNB test f1_score with best param: 0.470
Decision Trees	DecisionTreeClassifier train f1_score with best param: 0.731 DecisionTreeClassifier valid f1_score with best param: 0.709 DecisionTreeClassifier test f1_score with best param: 0.712
Logistic Regression	LogisticRegression train f1_score with best param: 0.823 LogisticRegression valid f1_score with best param: 0.799 LogisticRegression test f1_score with best param: 0.817
Linear SVM	LinearSVC train f1_score with best param: 0.907 LinearSVC valid f1_score with best param: 0.738 LinearSVC test f1_score with best param: 0.784

e) Based on the f1-score of the testing set with the best parameter: Logistic Regression > Linear SVM > Decision Trees > Bernoulli Naïve Bayes. And we can see that almost all classifiers achieved well except Bernoulli Naïve Bayes perhaps because the data is linearly separable, which will explain why **Logistic regression** and SVM worked so well, and because Bernoulli Naïve Bayes is a generative model I think I haven't tested the good range for the smoothing parameter alpha, that's why it didn't score well.

3- Using Medical-NLP dataset with FBoW representation and evaluating with the F1-score

a) I implemented Gaussian Naive Bayes, Decision Trees, Logistic regression and Linear SVM

b) The list of hyperparameters for each classifier:

Classifier	Hyperparameters and their range	Best values of hyperparameters
Gaussian Naïve Bayes		
Decision Trees	The criterion, tested gini and entropy The max depth of the tree, tested {10, 30, 50, 70} The minimum number of samples required to split an internal node, tested {0.1, 0.3, 0.5, 0.7, 0.9}	criterion: gini max_depth: 30 min_samples_split: 0.1
Logistic Regression	The norm of the penalty, tested l1 and l2 The regularization parameter, tested {0.01, 0.05, 0.1, 1, 5, 10, 50}	Penalty: l1 Regularization param C = 50
Linear SVM	The regularization parameter, tested the range [0.001, 0.1] with a step size 0.001	Regularization param C = 0.089

c) The training, validation, and test F1-score for all the classifiers

Classifier	F1-score with the best parameters
Gaussian Naïve Bayes	GaussianNB train f1_score: 0.690 GaussianNB valid f1_score: 0.364 GaussianNB test f1_score: 0.353
Decision Trees	DecisionTreeClassifier train f1_score with best param: 0.732 DecisionTreeClassifier valid f1_score with best param: 0.721 DecisionTreeClassifier test f1_score with best param: 0.709
Logistic Regression	LogisticRegression train f1_score with best param: 0.838 LogisticRegression valid f1_score with best param: 0.760 LogisticRegression test f1_score with best param: 0.765
Linear SVM	LinearSVC train f1_score with best param: 0.319 LinearSVC valid f1_score with best param: 0.317 LinearSVC test f1_score with best param: 0.312

d) Based on the f1-score of the testing set with the best parameter: Logistic Regression > Decision Trees > Gaussian Naïve Bayes > Linear SVM. And we can see that this time only 2 classifiers achieved well (Decision trees and Logistic regression), Gaussian Naïve Bayes didn't achieve well probably because the the assumption of a normal distribution is not a good assumption even when we take the frequency of each word, my guess is that the data this time is probably linearly separable (logistic regression working pretty well) but the classes do intersect a lot (SVM couldn't construct well his support vectors).

e) In terms of f1-score between BBOW and FBOW representation, we saw that we had 3 classifier achieving 0.7+ (and one achieving 0.8+) for BBOW and only 2 achieving 0.7+ for FBOW, so we can say that BBOW can be a better representation than FBOW (but with respect to the range of the hyperparameters I tested, different hyperparameters may have caused different scores). As for Naïve bayes, in BBOW we used Bernoulli Naïve bayes which is an event based model, while for FBOW we used Gaussian Naïve bayes which assumes the features are following a gaussian distribution, and we saw that both theses models weren't that good compared to other classifiers, perhaps because the assumptions of these 2 models do not apply well to our data.

f) In our case the BBOW representation is better, because it was easier for the models (especially the logistic regression) to separate linearly the different classes only by knowing which word exists in a transcript rather than knowing also the frequency of each one, the complexity of the information given by the frequency was not helpful in our case.