Naives Bayes and Logistic Regression

Data Sets	Data Representation	Training Algorithm	Accuracy	Precision	Recall	F1 Score
hw1	*Bernoulli*	Discrete Naïve Bayes	82.22	80.00	46.15	58.54
		Logistic Regression (λ=0.5, α=0.01, 1000)	**95.40**	93.08	**90.30**	91.67
		SGD Classifier	95.19	**99.23**	85.43	**91.81**
	Bag of Words	Multinomial Naïve Bayes	**96.86**	91.97	**96.92**	**94.38**
		Logistic Regression (λ=0.9, α=0.01, 1000)	93.72	90.00	87.31	88.64
		SGD Classifier	89.33	**98.45**	72.32	83.39
enron1	*Bernoulli*	Discrete Naïve Bayes	79.38	83.13	46.31	59.48
		Logistic Regression (λ=0.7, α=0.01, 1000)	94.74	93.29	**90.85**	92.05
		SGD Classifier	**94.96**	**97.99**	87.95	**92.70**
	Bag of Words	Multinomial Naïve Bayes	**94.96**	92.57	**91.95**	**92.25**
		Logistic Regression (λ=0.9, α=0.01, 1000)	94.74	92.62	91.39	92.00
		SGD Classifier	94.30	**95.97**	87.73	91.67
enron4	*Bernoulli*	Discrete Naïve Bayes	89.32	92.58	92.58	92.58
		Logistic Regression (λ=0.1, α=0.01, 1000)	96.50	**100.00**	95.36	97.63
		SGD Classifier	**96.87**	**100.00**	**95.83**	**97.87**
	Bag of Words	Multinomial Naïve Bayes	88.21	96.85	86.44	91.35
		Logistic Regression (λ=0.7, α=0.01, 1000)	**95.39**	**99.74**	**94.20**	**96.89**
		SGD Classifier	94.66	99.49	93.51	96.41

The hyper-parameters Lambda, Learning Rate, and the number of iterations. The Lambda is assumed to be in the range of 0.1 to 0.9 with a difference of 0.1 and fixed iterations and learning rate and tested it on the validation data (part of the full train data). The final Lambda value which gives highest accuracy is set as regularization value for the testing purpose.

The hard limit on the number of iterations is set to 1000, even for the higher number of iterations the gradient decreases and difference in the weights is very less and almost same. For this hard limit of iterations, the learning rate should be set to a smaller or little higher value and, the weights are almost close to the converging point. We obtain maximum or minimum where the slope reaches zero and thus, we set the learning rate to 0.01. This results somewhat faster convergence. If the learning rate is set to a larger value it will overshoot the minima and oscillates.

The following are the answers to the assignment questions.

1. Bernoulli Dataset representation along with SGD Classifier gives better performance results when compared to other algorithms considering all the four metrics (precision, accuracy, recall, F1 score). SGD Classifier is performing almost same for Bernoulli and Bag of Words data representation. Because for some word frequencies weights factor will be almost same as Bernoulli model. And, in some cases LR perform better than SGD Classifier. When compared to both naïve bayes Multinomial and Discrete the LR and SGD Classifier are far better.

2. Yes, sometimes Multinomial Naïve Bayes perform better than LR and SGD Classifier on the Bag of Words representation. In these cases, the Naïve Bayes assumptions might be correct. If both the precision and recall are better than the F1 Scores, then this implies prediction of spam as ham is less and also prediction of ham as spam is minimal. In LR and SGD Classifier some weights might become zero even though word frequencies are not and this eliminate that feature in LR and SGDC whereas in Multinomial these features might not be eliminated.

3. No, the Discrete Naive Bayes doesn't perform better than the LR and SGD Classifier on the Bernoulli Representation. The representation is same in all the three cases but we also consider non-occurrences in Discrete Naïve Bayes which might decrease our probability, whereas in LR and SGDC weights are not allotted to non-occurrences.

4. In most of the cases the LR and SGD Classifier perform almost same and, in some cases, it outperformed the SGD Classifier. Because in such case the convergence might be more efficient in LR so produced good results.