BITS F464 - ML Assignment -2

Logistic Regression and Naive Bayes

Group Members:

1	**N Vishnuteja Reddy**	**2016A7PS0086H**
2	**Sanjay D**	**2016A7PS0033H**
3	**Athilesh Arputham**	**2016A7PS0139H**

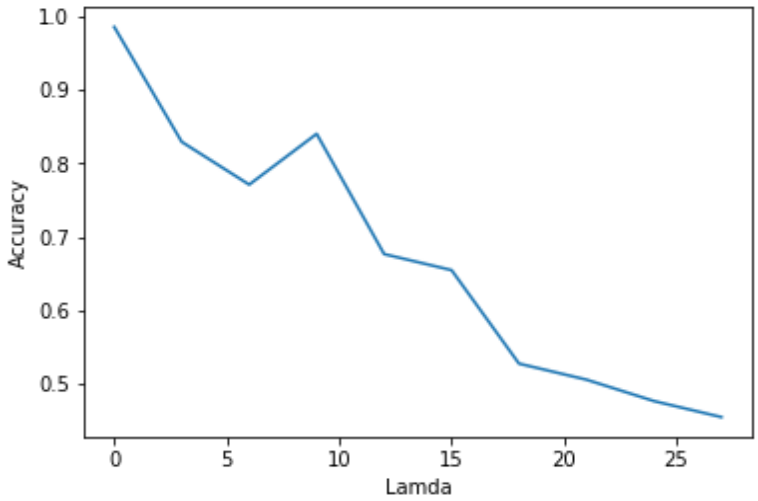
Logistic Regression

The following observations were made when logistic regression was run without regularisation.

Learning Rate	Initialisation	Accuracy	Iterations	Weights
0.075	[1,1,1,1,1]	0.989	4973	[7.30083571, -8.03246109, -4.23436218, -5.35180368 -0.57270676]
0.075	random	0.978	6509	[7.17622659, -7.68794145, -4.13535894, -5.17930533 -0.62604835]
0.01	[1,1,1,1,1]	0.989	38160	[6.40415737, -6.93971584, -3.72102653, -4.68569591, -0.53317087]
0.01	random	0.978	32072	[6.86723098 ,-6.89235873, -3.84801378, -4.76699635, -0.60917061]
0.001	[1,1,1,1,1]	0.989	100000	[7.40307225, -7.27813593, -4.08010389, -5.09872525, -0.62382601]
0.001	random	0.985	100000	[6.57181931, -7.35047519, -3.89590984, -4.87780828, -0.50314081]

With Regularisation:

Regularisation parameter	Weights
0	**[7.30083571, -8.03246109, -4.23436218, -5.35180368, -0.57270676]**
2	**[5.74586075, -7.27257821, -2.54476173, -1.81847524, 0.54424843]**
4	**[9.66599197, -4.66232475, -3.08879005, 0.32914523, 3.68173986]**
6	**[3.84358591, -3.48889682, -1.7773333, -1.01709102, 2.75254866]**
8	**[8.94786844, -3.41106925, -2.64451708, 0.63934405, 5.02107887]**
10	**[14.34158716, -5.78888581, -4.941005, 4.05203679, 8.27161158]**
12	**[-10.43264173, -5.86210934, 1.26450172, -3.1831026, -7.03513068]**
14	**[-15.82995897, -5.23624246, 2.10292263, -3.95275429, -8.72173748]**
16	**[-22.1915492, -4.58984966, 3.11371647, -4.99152297, -10.74458752]**
18	**[-29.78448341, -4.29079258, 4.46823515, -6.4700196, -13.6797834]**



When regularisation is used, it is found that the model has a better accuracy without any kind of regularisation. The regularised parameter varies from 0 to 25 and it is plotted against accuracy as shown.

Naive Bayes

Confusion Matrix

	Predicted Class pos	**Predicted Class neg**
Actual Class pos	1012	141
Actual Class neg	210	1020

Naive Bayes Model	**Precision**	**Recall**	**F-Score**	**Accuracy**
Pos	82.81	87.77	85.22	85.27
Neg	87.85	82.926	85.319	NA

Remarks:-

Naive Bayes provides satisfactory results for sentiment analysis in text classification.

The probability of occurrence of any word give class label is independent of the probability of any other word given that label.

The probability of occurrence of any word in a document is independent of the location of the word in the document.