Programming Assignment-01
Intro to Machine Learning -CS5011

Sunil Kumar J S(NA13B031)

September 20, 2016

1

1.1 Details of parameter

Mean vector:
Mean vectors are generated by randomly gnerating numbers in range 0 to 2.By fixing range 0 to 2.This range was chosen using trail and error method so that there is some ovelap in the data

Covariance Matrix:
Generation of common covariance matrix: common covariance matrix has to be symmetric and postive definite.To ensure this, covariance matrix is generated in following steps:

- Generated a random matrix of size (20,20) in which elements are in range (0 ,1)
- obtained new matrix ,B= 0.5(A+ A')
- covarince matrix is B.B'

1.2 LDA performance

LDA is fitted for the generated data. Performance Results obtained are as follows

Metric	LDA	
	class 1	class 2
Precision	0.993	0.998
Recall	0.998	0.993
f-measure	0.995	0.995

Table 1: Peformace results of LDA

2

2.1 Best fit resullts for linear Regressors:

Metric	Linear Regression	
	class 1	class 2
Precision	0.993	0.998
Recall	0.998	0.993
f-measure	0.995	0.995
accuracy	0.995	

Table 2: Peformace results of Linear regression classifier

2.2 Coefficients learnt:

$$\beta = \begin{pmatrix} -0.16, & -0.15, & -0.029, & -0.20, & 0.49, & 0.28, & -0.36, & -0.09, & -0.009, & 0.33 & \\ & 0.26, & -0.24, & -0.51, & 0.011, & 0.11, & 0.51, & 0.035, & 0.16, & -0.115, & -0.28 \end{pmatrix}$$

*β is row vector, due to space constraint it is represented in two rows
intercept=0.602

3

3.1 Performance k-NN

k-NN classifier didn't perform better than Linear regression classifier. For all values of k performance of k-NN is worse than linear regression classifier

3.2 k-NN Best fit results

k-NN performed best when k=61. The best fit metrics for different values of k are as follow

Metric	k-NN	
	class 1	class 2
Precision	0.822	0.836
Recall	0.839	0.818
f-measure	0.830	0.827
accuracy	0.829	

Table 3: Peformace results of LDA

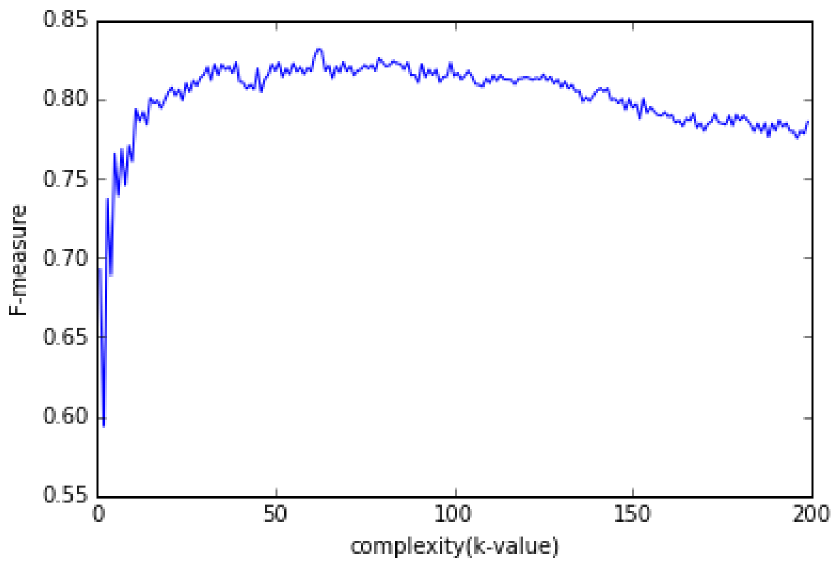


Figure 1: F-measure of k-NN classifier with varying k

4

4.1 Is sample mean good choice?

No.mean imputation attenuates any correlations involving the variable(s) that are imputed. This is because, imputed values of a attribute solely depends upon that particular attribute. However in the dataset it's intuitive that attributes have correlation.

4.2 Other Method

Multiple Imputation methods are shown to work better as it simulate both the process generatating the data and the uncertainity associated with the parameters of the probability distribution of the data. Also missing values in given data follows **Monotone pattern**. Therefore **Propensic Score Method** is adopted.
Steps for Impuation by Propensic Score Method:

- Create an indicator variable Rj with the value 0 for observations with missing Yj and 1 otherwise.
- Fit a logistic regression model of for indicator Rj on complete columns.
- Create a propensity score for each observation to indicate the probability of its being missing
- Divide the observations into a fixed number of groups based on these propensity scores.
- Apply an approximate Bayesian bootstrap imputation toeach group.

Referencehttp://www.ats.ucla.edu/stat/sas/library/multipleimputation.pdf

4.3 completed dataset is named as completed.csv in datasets folder

5

Residual error for best fit is 7.52
Coefficients learnt are mentioned in coefficients-5.txt file attached

6

6.1 λ value for best fit is 39

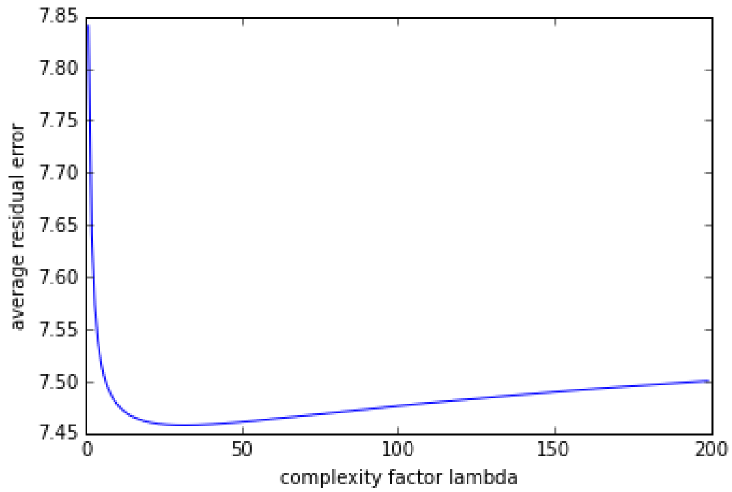


Figure 2: Average residual error against lambda

λ	Avg Residual Error
10	7.81
20	7.38
39	7.377
50	7.378
2000	7.49

Table 4: Average Residual Error for different values of λ

6.2 Best fit for reduced data set

Residual error of the best fit it 7.27

6.3 Feature Selection

Yes. Ridge regression could be used as feature selection by selecting only those features which have regression coefficients greater than certain threshold.From trial and error method it is found that forcing coefficients(absolute value) less than **0.028** to zero yielded best performance

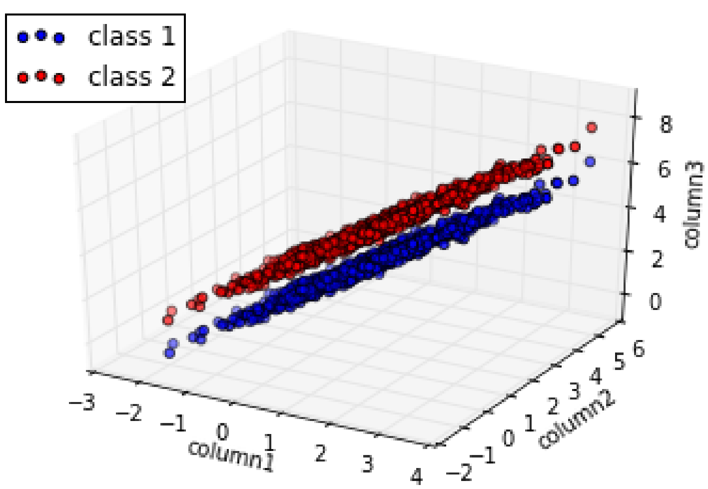
7

7.1 Performance of PCA

Metric	PCA	
	class 1	class 2
Precision	0.612	0.613
Recall	0.615	0.610
f-measure	0.613	0.612

Table 5: Peformace results of PCA

7.2 3-D plot of dataset



7.3 Plot of the dataset in the projected space

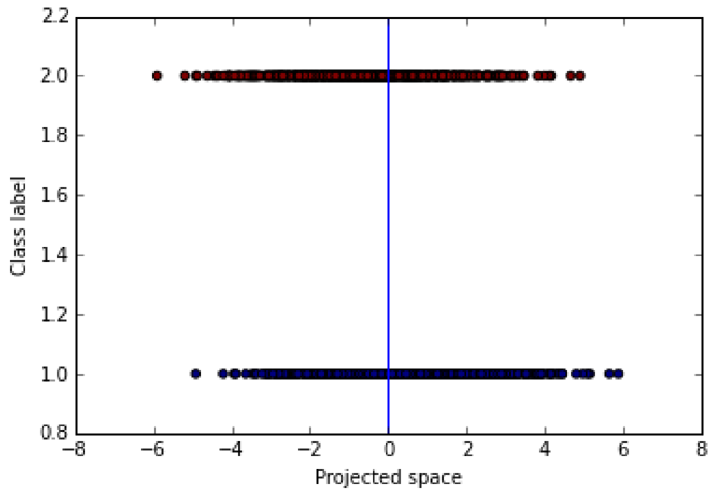


Figure 3: *while plotting height equal to class label is given so that we could differentiate classes

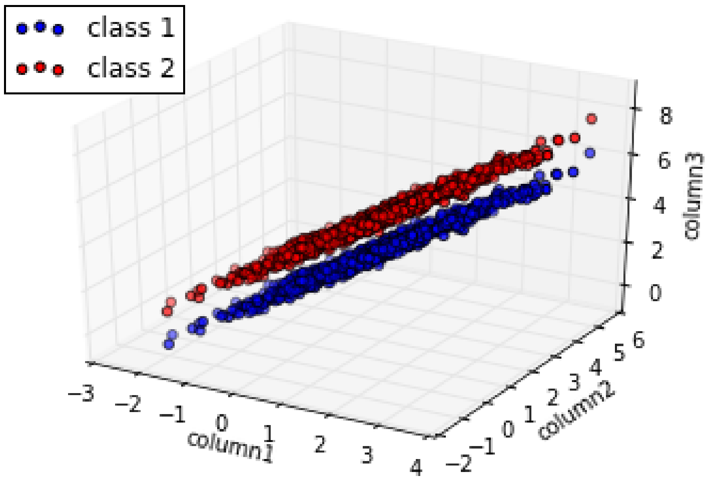
8

8.1 Performance of LDA

Metric	PCA	
	class 1	class 2
Precision	1.0	1.0
Recall	1.0	1.0
f-measure	1.0	1.0

Table 6: Peformace results of LDA

8.2 3-D plot of dataset



8.3

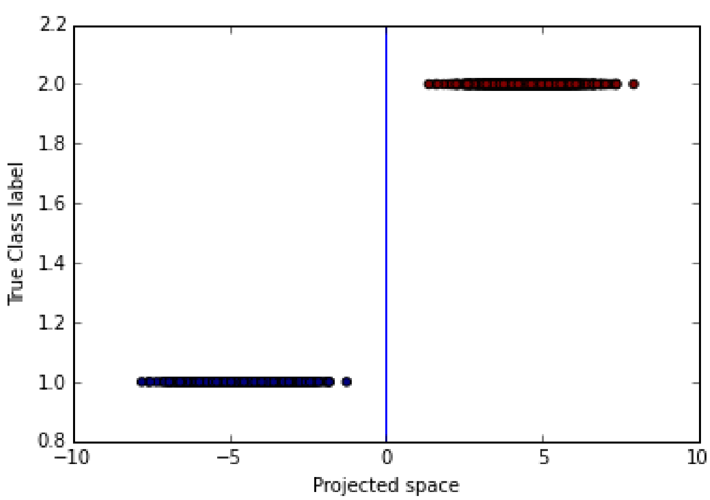


Figure 4: *while plotting height equal to class label is given so that we could differentiate classes

8.4 Inference

- LDA performs better than PCA. This is because LDA finds the direction which maximizes the **between class variance** relative to **within class variance** however PCA finds direction along which **entire data** could be projected with maximum variance.

9

9.1 Logistic Regression performace results

Metric	LR	
	mountain	forest
Precision	0.933	0.760
Recall	0.699	0.949
f-measure	0.800	0.844

Table 7: Peformace results of LR

9.2 L regularized LR performace results

Metric	L1 LR	
	mountain	forest
Precision	1.0	1.0
Recall	1.0	1.0
f-measure	1.0	1.0

Table 8: Peformace results of L1 LR

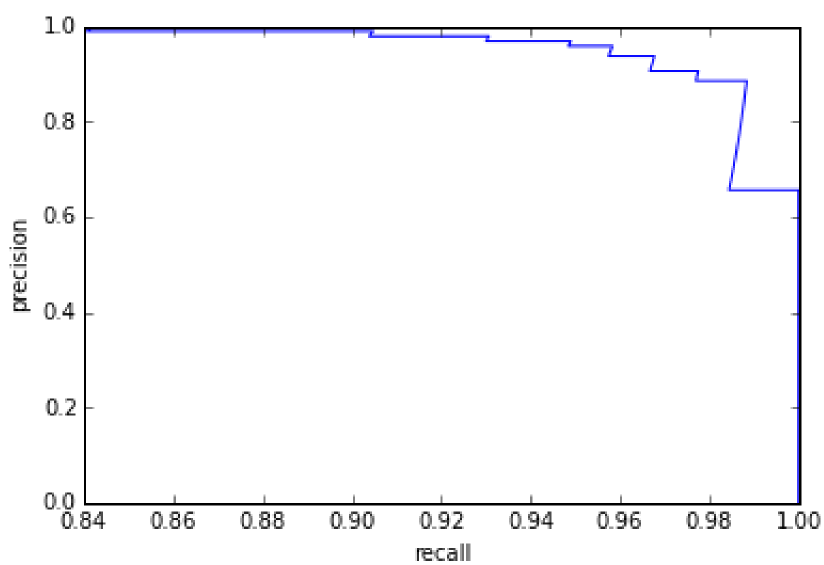
10

10.1 Maximum Likelihood asssuming Likelihood is Multinomial

Metric	Class spam
Precision	0.930
recall	0.969
F-measure	0.949

Table 9: Peformace results of Multinomial

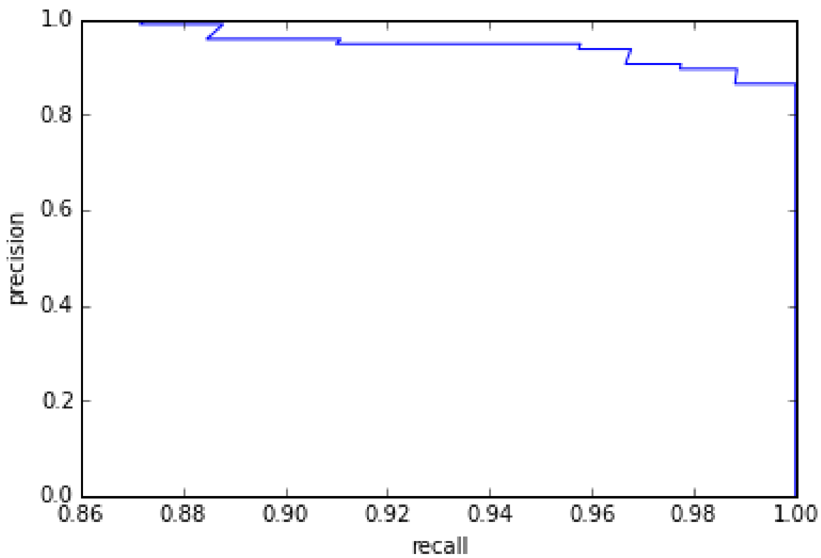
Please note that limits of x axis is changing with plots so that precision plot could be analysed better in critical regions.For all values less than minimum recall,precision is 1.AUC criteria can't be applied for following graphs



10.2 Maximum Likelihood asssuming Likelihood is Bernouli

Metric	Class spam
Precision	1.0
recall	0.844
F-measure	0.915

Table 10: Peformace results of Multinomial



10.3 Maximum Likelihood asssuming Likelihood is Multinomial with Dirchlet prior

Metric	Class spam
Precision	0.990
recall	1.0
F-measure	0.995

Table 11: Peformace results of Multinomial

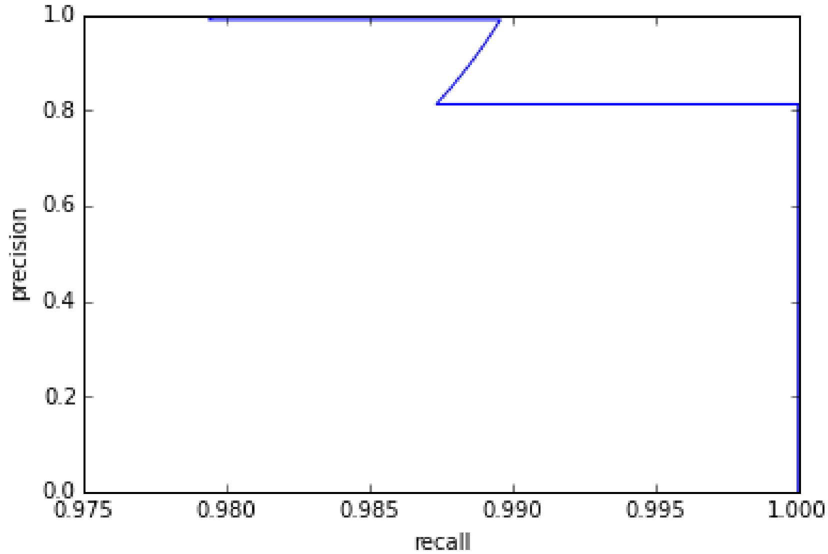


Figure 5: alphas(hyperparameters) are taken to be equal to 0.01 in this graph

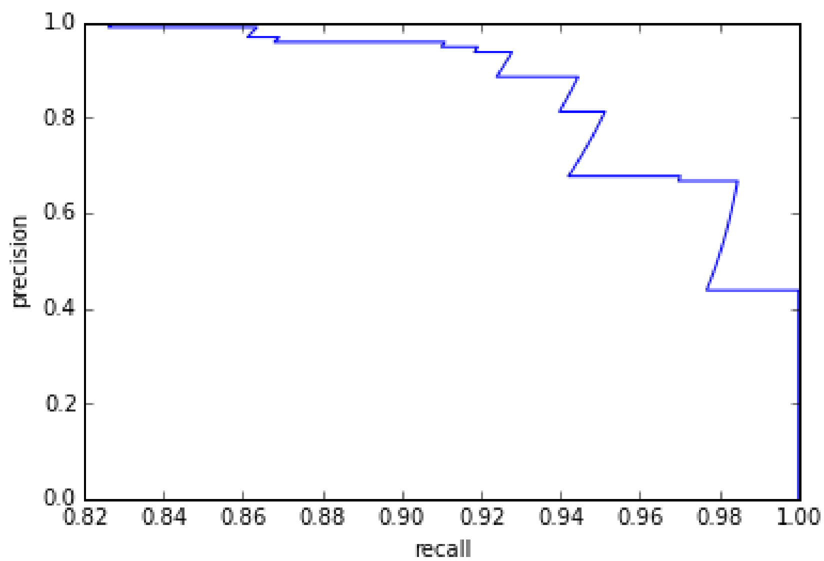


Figure 6: alphas(hyperparameters) are taken to be equal to 10 in this graph

10.4 Maximum Likelihood asssuming Likelihood is bernouli with Beta prior

Metric	Class spam
Precision	0.989
recall	0.958
F-measure	0.974

Table 12: Peformace results of Beta prior

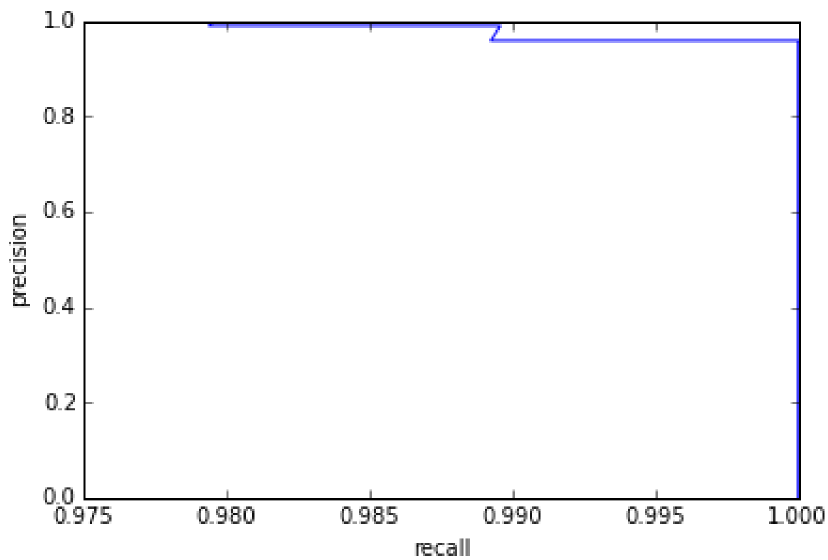


Figure 7: alpha and beta(hyperparameters) are taken to be equal to 0.0001 in this graph

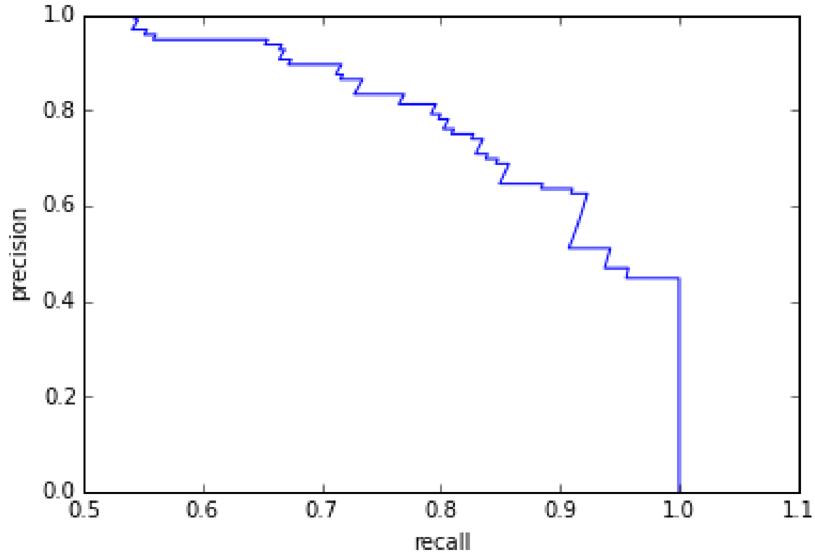


Figure 8: alphas(hyperparameters)are taken to be equal to 10 in this graph

10.5 comments

test set	Multinomial	Bernouli	Multinomial(with Dirchlet prior)	Bernouli with Beta (prior)
(1,2)	0.955	0.945	0.990	0.984
(3,4)	0.969	0.917	0.968	0.957
(5,6)	0.960	0.867	0.979	0.974
(7,8)	0.959	0.921	0.989	0.957
(9,10)	0.938	0.871	0.968	0.957

Table 13: F-measure for different models for different test sets

- As shown in above fig,choice of good priors yeilded better performance in each case
- In maximum likelihood estimation, Multinomial outperforms Bernouli because Bernouli in Bernouli we just keep track if word is present in document or not but in Multinomial we also account for number of times of occurances of word

- Choice of prior (α) impacts model performance in Bayesian estimation.For example when alphas (hyperparameters for Dirchlet prior) are taken to be 0.01 Bayesian outperforms maximum likelihood however when alphas are taken to as 10 ,Bayesian performs worse than maximum likelihood as it needs more data to learn