1. Problem

Finished follows problems by the algorithm in class

① 1D 2-class Gaussian discriminant analysis
② nD 2-class Gaussian discriminant analysis
③ nD k-class Gaussian discriminant analysis
④ Naive Bayes with Bernoulli features
⑤ Naive Bayes with Binomioal features

2. Proposed Solution

① 1D 2-class Gaussian discriminant analysis

Get model parameters using follow equation:

$$u = \frac{1}{m}\sum x^{(i)}$$

$$\sum = \frac{1}{m}(x^{(i)} - u)(x^{(i)} - u)^T$$

Discriminant Function:

$$g_i(x) = P(y = i \mid x) = \log(\frac{1}{2\pi}) - \log(\sigma_i) - \frac{1}{2}\frac{(x - u_i)^2}{\sigma_i^2} + \log(p \mid y = i)$$

Confusion Matrix

	P	N
P	TP	FP
N	FN	TN

Precision = $\frac{TP}{TP + FP}$

Recall = $\frac{TP}{TP + FN}$

Accuracy = $\frac{TP + TN}{TP + FP + FN + TN}$

② nD K-class(2-class)Gaussian discriminant analysis

$$m_i = \sum_{i=1}^{m} 1(y^{(j)} = i)$$

$$u_i = \frac{1}{m}\sum_{i=1}^{m} 1(y^{(j)} = i)x^{(i)}$$

$$\sum_i = \frac{1}{m_i}\sum_{i=1}^{m} 1(y^{(j)} = i)(x^{(i)} - u_i)(x^{(i)} - u_i)^T$$

Discriminant Function:

$$g_i(x) = P(y = i \mid x) = \log(\frac{1}{2\pi}) - \log(\sigma_i) - \frac{1}{2}\frac{(x - u_i)^2}{\sigma_i^2} + \log(p \mid y = i)$$

Precision

	C1	C2	C3	…	Cn
C1					
C2					
C3					
. . .					
Cn					

Precision (Ci)= $\frac{c_{ii}}{\sum_{ij} c_{ij}}$

Recall (Ci)= $\frac{c_{ii}}{\sum_{j} c_{ij}}$

F-Measure = $2\frac{precisson \times recall}{precission + recall}$

③ Naive Bayes

Classification

Find :p(y=j|x)

X1 X2 X3 … Xm	Y1
X_{m+1} X_{m+2} … X_k	Y2
…	…

Bayes rule

$$p(y=j \mid x)=\frac{p(y=j,x)}{p(x)}=\frac{p(x \mid y=j)p(y=j)}{p(x)}=\frac{p(x \mid y=j)p(y=j)}{\sum p(x \mid y=j)p(y=j)}$$

3. Implementation Detail

For Gaussian Discriminant

① Read data

```
Iris <- iris[1:100,4:5]
Iris <- droplevels(Iris)
```

② using linear Discriminant Analysis get model

```
model <- lda(Species ~ ., data = Iris)
```

③ get prediction

```
pred    <- predict(model, Iris[,1:2])
```

yhat=apply(pred$posterior,1,which.max)

④ get confusion matrix

table(yhat, Iris[,2])

⑤ Get precision recall

pred <- prediction(pred$posterior[,1],y)

perf <- performance(pred, "prec","rec")

plot(perf, xlim = c(0,1), ylim = c(0,1))

⑥ Get F-measure

fmeasure <- performance(pred,"f")

plot(fmeasure)

⑦ Get Accuracy

accuracy <- performance(pred,"acc")

plot(accuracy)

⑧ Cross Validation

For Naive Bayes with Bernoulli feature

① Read Data

SPEC <- read.table("F:/SPEC.data",sep=",")# 23 columns

SPEC <- droplevels(SPEC)

② Get Bayes model

model<-naiveBayes(SPEC[,1:22], SPEC[,23])

③ Get Prediction

yhat = predict(model, SPEC[,-23])

④ Get precision recall

yhat = predict(model, SPEC[,-23],type="raw")

pred <- prediction(yhat[,1],y)

perf <- performance(pred, "prec","rec")

⑤ Get f-measure

fmeasure <- performance(pred,"f")

plot(fmeasure)

⑥ Get accuracy

accuracy <- performance(pred,"acc")

For Naive Bayes with Binomioal feature

① Read Data

Toe <- read.table("F:/TOE.data",sep=",")# 23 columns

Toe <- droplevels(Toe)Get Bayes model

② Get naiveBays model

model<-naiveBayes(Toe[,1:9], Toe[,10])

③ Get Prediction

yhat = predict(model, Toe[,-10])

④ Get precision recall

yhat = predict(model, Toe[,-10],type="raw")

pred <- prediction(yhat[,1],y)

perf <- performance(pred, "prec","rec")

⑤ Get f-measure

fmeasure <- performance(pred,"f")

plot(fmeasure)

⑥ Get accuracy

accuracy <- performance(pred,"acc")

4. Results and discussion

1D 2-Class Gaussian discriminant analysis

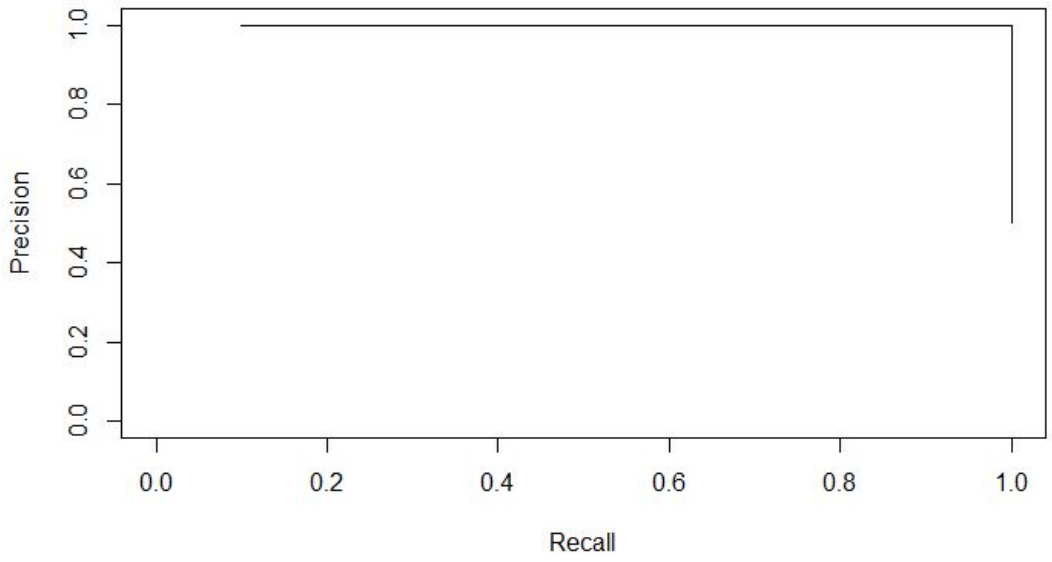
Confusion Matrix

yhat	setosa	versicolor
1	50	0
2	0	50

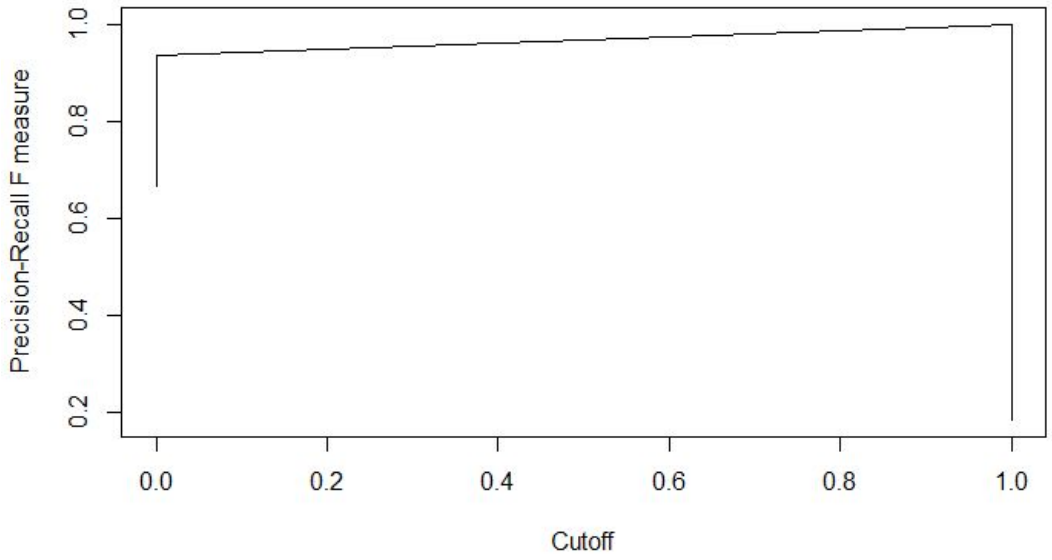
Cross Validation

yhat	setosa	versicolor
1	50	0
2	0	50

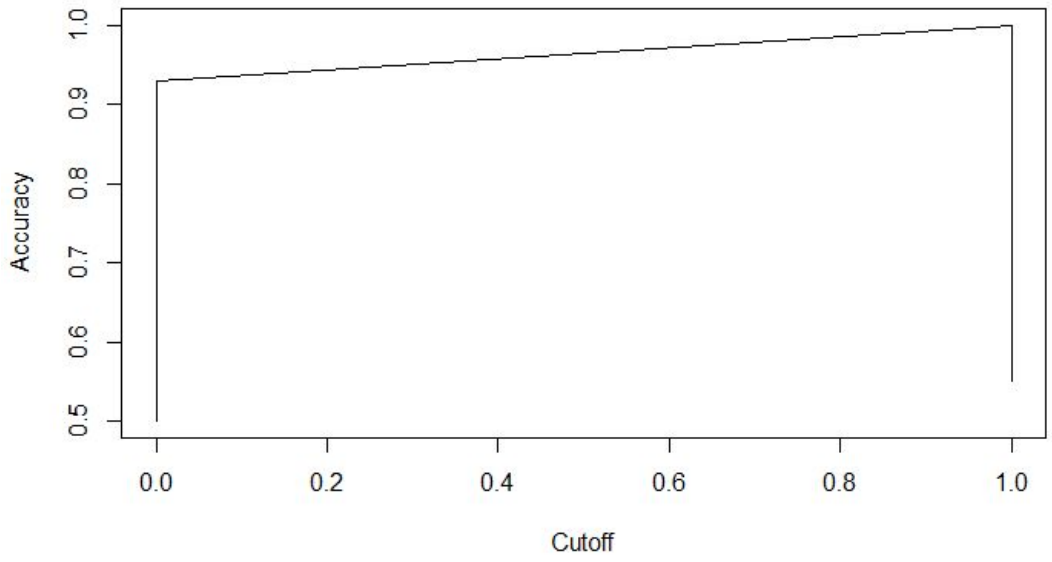
Precision , recall



F-measure



Accuracy



n D 2-Class Gaussian discriminant analysis

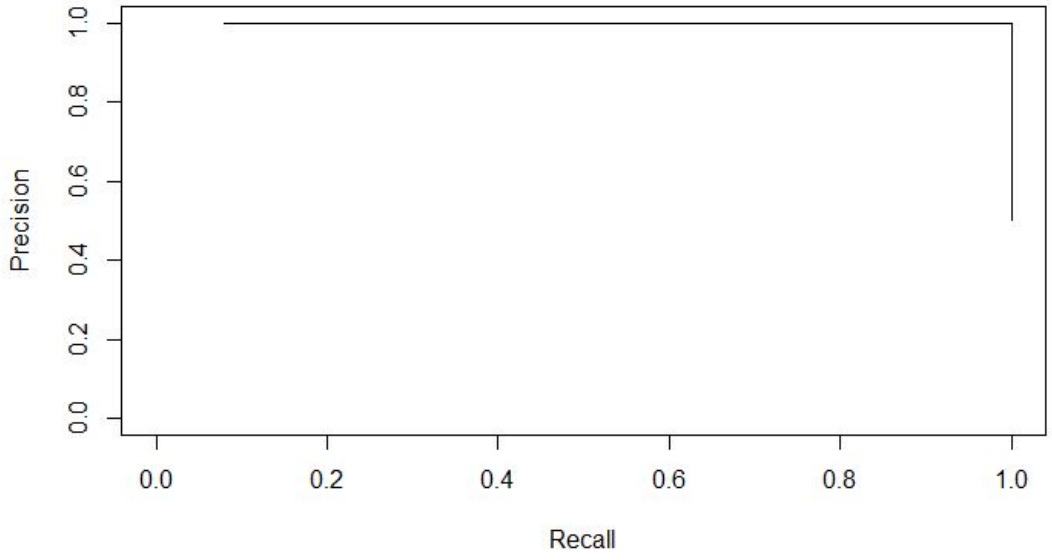
Confusion matrix

yhat	**setosa**	**versicolor**
1	**50**	**0**
2	**0**	**50**

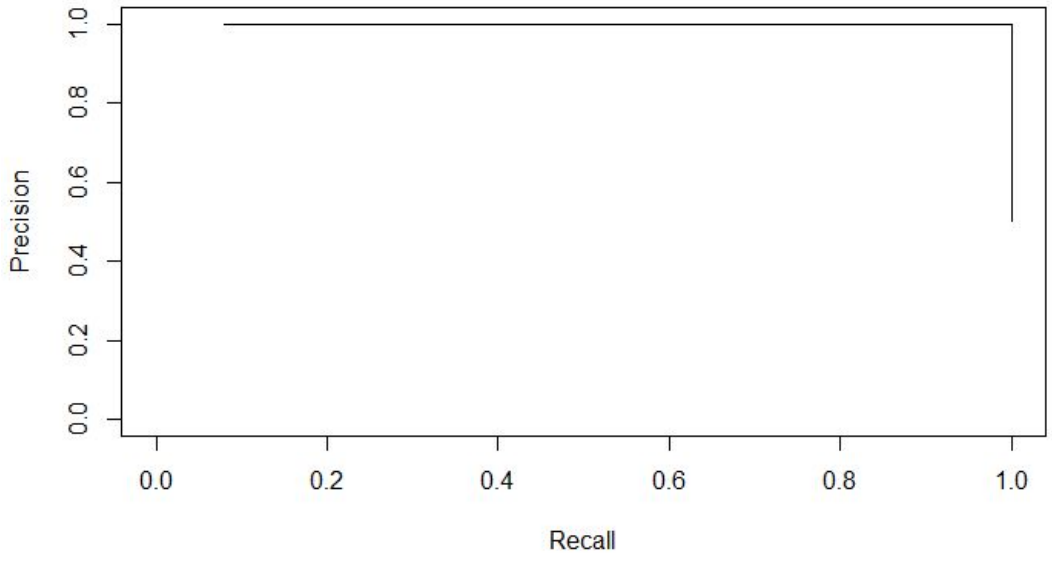
Cross Validation

yhat	**setosa**	**versicolor**
1	**50**	**0**
2	**0**	**50**

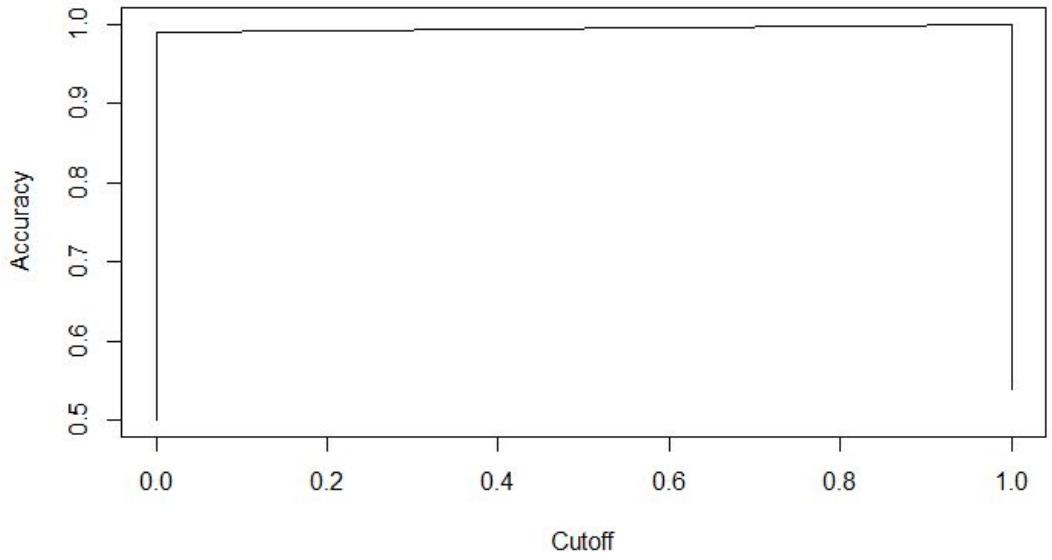
Precision , Recall



F-measure



Accuracy



n D K-Class Gaussian discriminant analysis

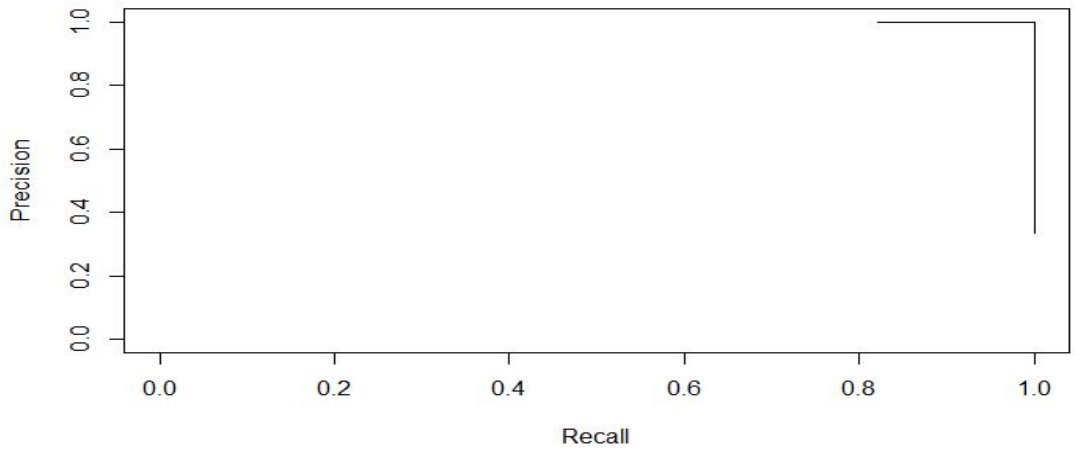
Confusion matrix

yhat	setosa	versicolor	virginica
1	50	0	0
2	0	48	1
3	0	2	49

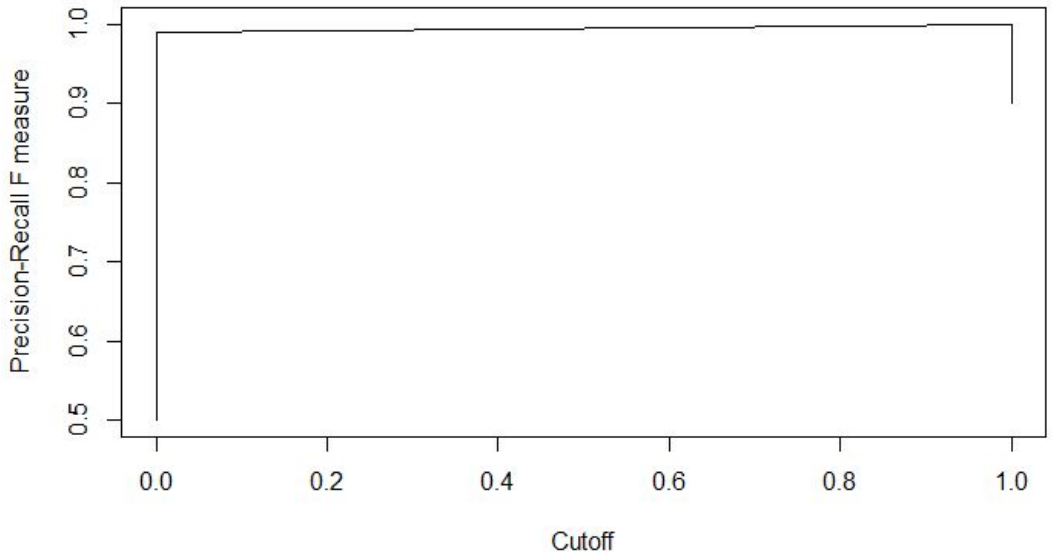
Cross Validation

yhat	setosa	versicolor	virginica
1	50	0	0
2	0	48	1
3	0	2	49

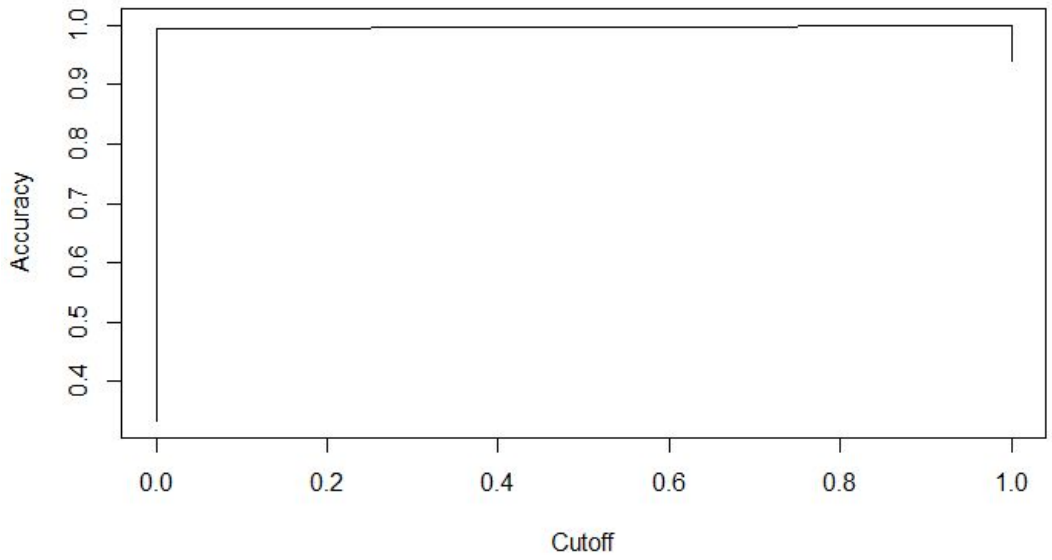
Precision , Recall



F-measure



accuracy



Naive Bayes with Bernoulli features(2-class)

Confusion matrix:

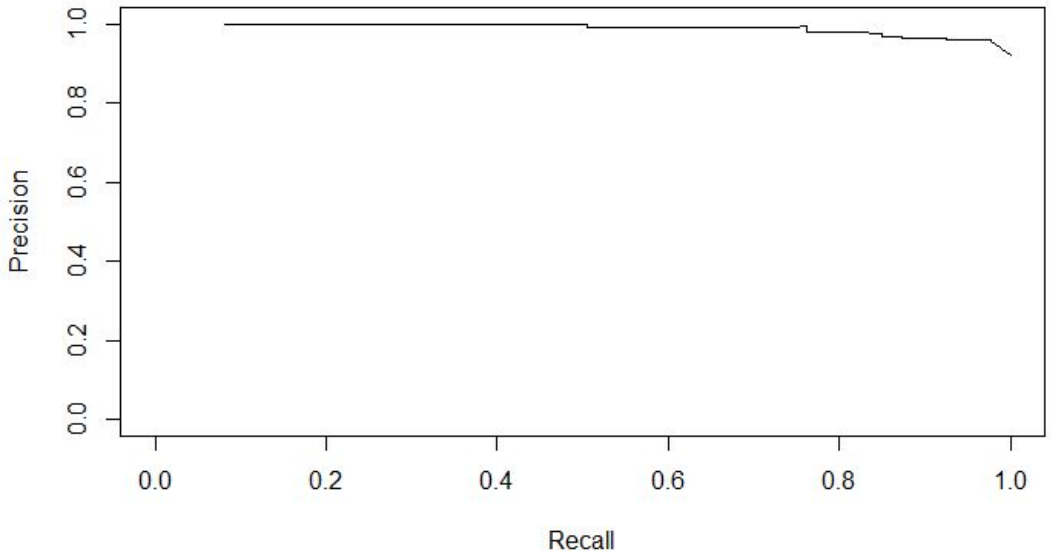
yhat	one	zero
one	87	1
zero	85	4

Cross Validation

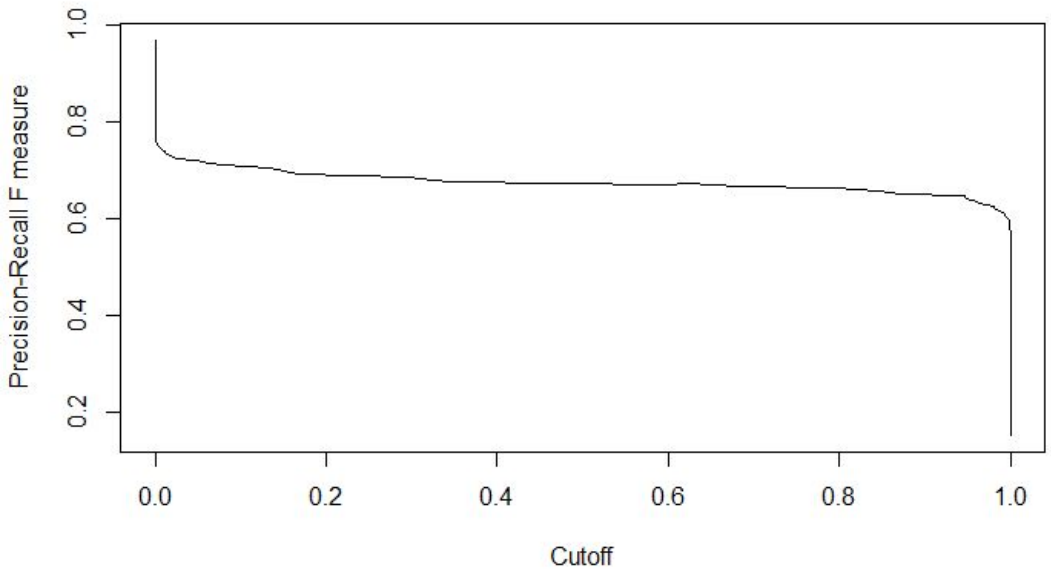
yhat	one	zero
one	70	1

zero	102	14

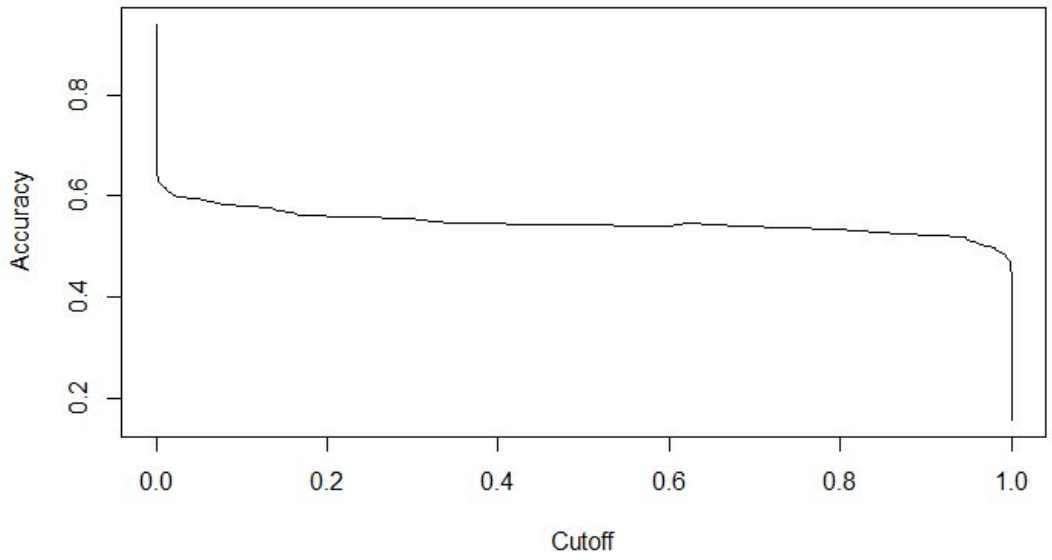
Precision-recall



F-measure



Accuracy



Naive Bayes with Binomioal features(2-class)

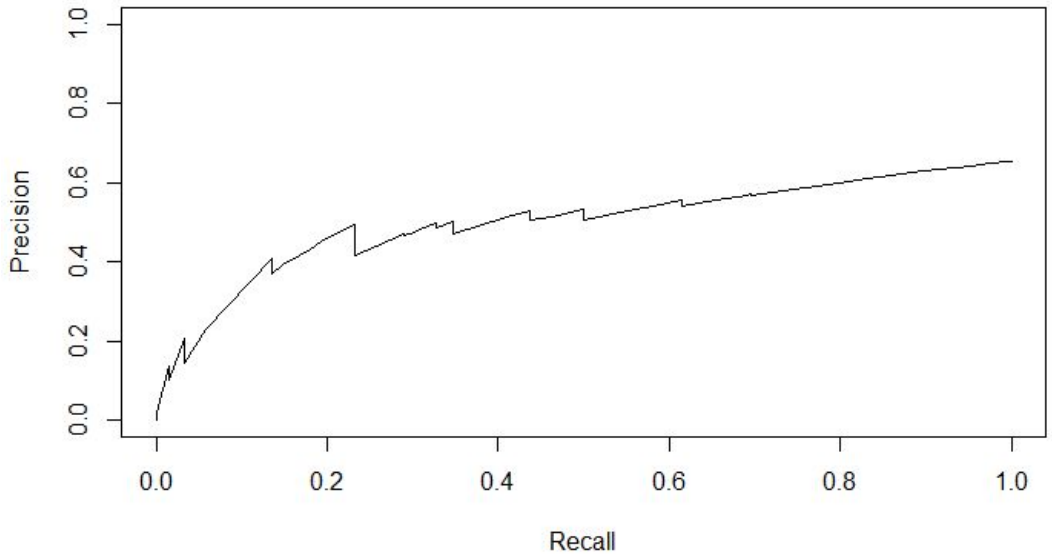
Confusion Matrix

yhat	negative	positive
negative	148	105
positive	184	521

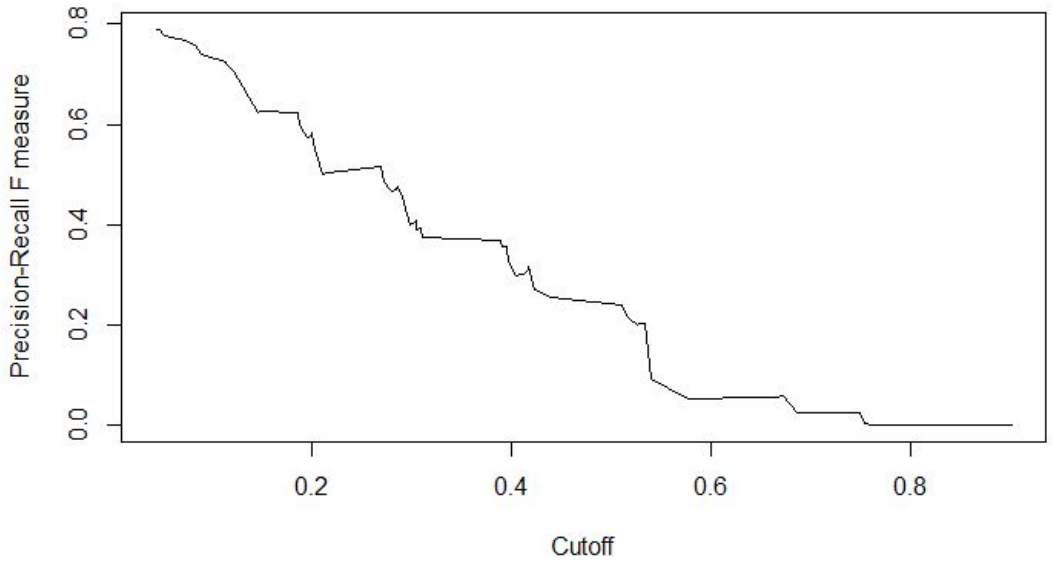
Cross Validation

yhat	negative	positive
negative	142	91
positive	190	535

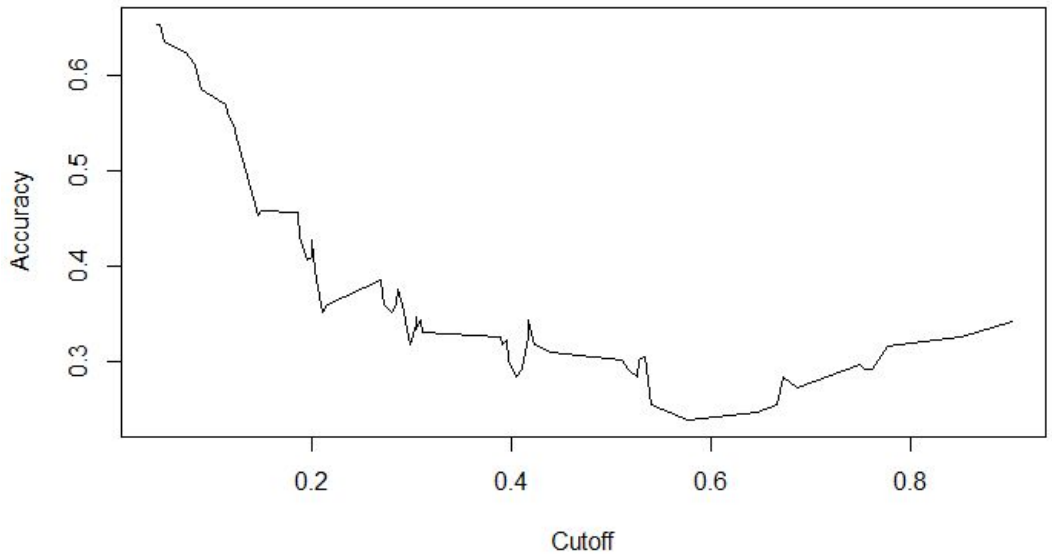
Precision,recall



F-measure



Accuracy



5. Reference

Elements of Statistical Learning (Second Edition)

http://www-users.cs.york.ac.uk/~jc/teaching/arin/R_practical/

http://stat.ethz.ch/R-manual/R-patched/library/MASS/html/lda.html

http://lausanne.isb-sib.ch/~darlene/gda/tp/tp01.html

http://cran.r-project.org/web/packages/cvTools/cvTools.pdf

http://cran.r-project.org/web/packages/ROCR/ROCR.pdf