CIS – 490: Machine Learning
Learning Activity 1
Name: Pranav Vinod
UMassD ID: 01984464

Naïve Bayes

1. Split data into training and testing in 50:50 ratio

```
library(e1071)
library(pROC)
data(iris)

set.seed(490)

# train-test split
training.indices <- sample(1:nrow(iris), 0.5 * nrow(iris))
x.train <- iris[training.indices, 1:4]
x.test <- iris[-training.indices, 1:4]
y.train <- iris[training.indices, 5]
y.test <- iris[-training.indices, 5]
```

2. Construct the model

```
# Construct the model
model <- naiveBayes(x.train, y.train)
# Predict for training
pred.train <- predict(model, x.train, type="class")
# pred.train
conf.train <- table(pred.train, y.train)
conf.train
```

3. Make predictions

```
# Predict for testing

# type=class tells model to output specific class predicted (setosa, versicolor, or # virginic
pred <- predict(model, x.test, type="class")
conf <- table(pred, y.test)
conf

# Predict for all data

pred.all <- predict(model, iris[1:4], type="class")
conf.all <- table(pred.all, iris$Species)
conf.all
```

4. Calculate Accuracy

```r
# Train set
acc.train <- sum(diag(conf.train)) / sum(conf.train)
acc.train
err.train <- 1 - acc.train
err.train

# Test set
acc <- sum(diag(conf)) / sum(conf)
acc
err <- 1 - acc
err

# All data
acc.all <- sum(diag(conf.all)) / sum(conf.all)
acc.all
err.all <- 1 - acc.all
err.all
```

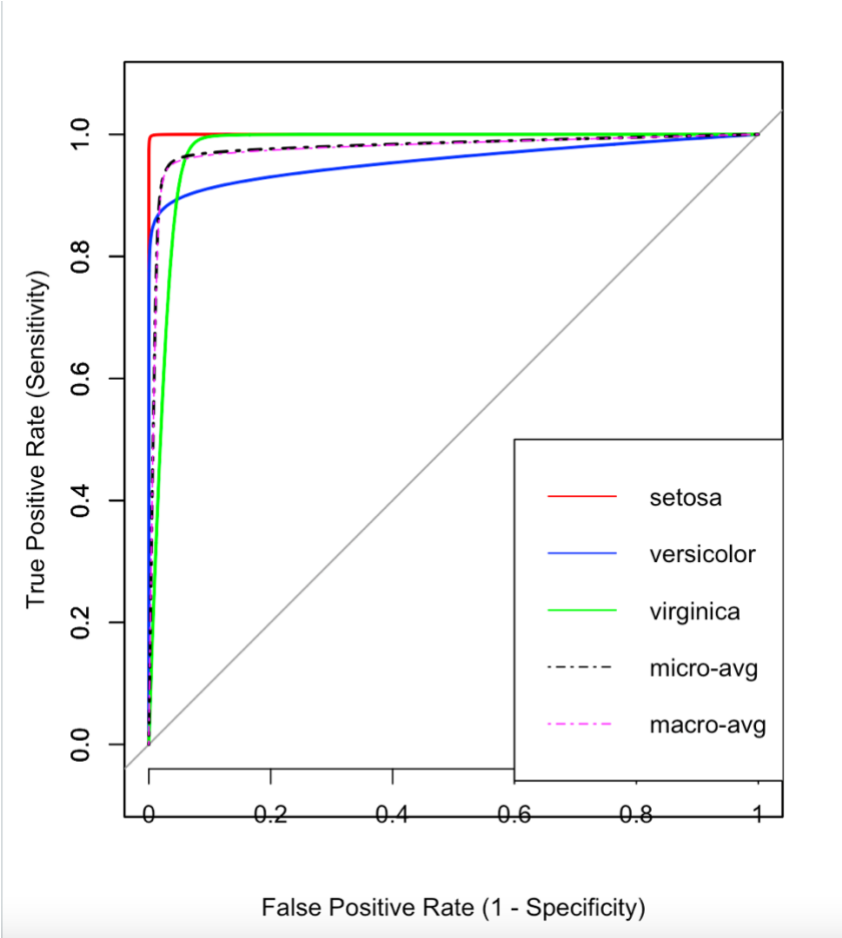
5. Model Evaluation

		Train	Test	All
Accuracy		0.95	0.95	0.95
AUC		0.996	0.997	0.995
Macro Sensitivity		0.95	0.95	0.95
Macro Specificity		0.97	0.97	0.97
Sensitivity	Setosa	1.00	1.00	1.00
	Versicolor	0.96	0.85	0.90
	Virginica	0.89	1.00	0.94
Specificity	Setosa	1.00	1.00	1.00
	Versicolor	0.94	1.00	0.97
	Virginica	0.98	0.92	0.95
PPV	Setosa	1.00	1.00	1.00
	Versicolor	0.88	1.00	0.94
	Virginica	0.96	0.85	0.90
NPV	Setosa	1.00	1.00	1.00
	Versicolor	0.98	0.92	0.95
	Virginica	0.94	1.00	0.97

6. ROC Curves

```r
# ROC curve using Test Set

probs <- predict(model, x.test, type="raw")

setosa.labels <- rep(0, length(y.test))
versicolor.labels <- rep(0, length(y.test))
virginica.labels <- rep(0, length(y.test))
for (i in 1:length(y.test)) {
  if(y.test[i] == 'setosa') {
    setosa.labels[i] <- 1
  } else if (y.test[i] == 'versicolor') {
    versicolor.labels[i] <- 1
  } else if (y.test[i] == 'virginica') {
    virginica.labels[i] <- 1
  }
}

# Create a plot of ROC for each class label against the rest
setosa.roc <- roc(setosa.labels, probs[, 'setosa'], auc.polygon=TRUE, max.auc.polygon=TRUE, pr
setosa.smoothroc <- smooth(setosa.roc, method = "density")
plot(setosa.smoothroc, col = 'red', xaxt='n', xlab="False Positive Rate (1 - Specificity)", yla

#par(new=T): make the second plot without cleaning the first
par(new=TRUE)
versicolor.roc <- roc(versicolor.labels, probs[, 'versicolor'], auc.polygon=TRUE, max.auc.polyg
versicolor.smoothroc <- smooth(versicolor.roc, method = "density")
plot(versicolor.smoothroc, col='blue', xaxt='n', xlab="", ylab = "")

par(new=TRUE)
virginica.roc <- roc(virginica.labels, probs[, 'virginica'], auc.polygon=TRUE, max.auc.polygon=
virginica.smoothroc <- smooth(virginica.roc, method = "density")
plot(virginica.smoothroc, col='green', xaxt='n', xlab="", ylab = "")

y.labels <- c(setosa.labels, versicolor.labels, virginica.labels)
y.probs <- c(probs[, 'setosa'], probs[, 'versicolor'], probs[, 'virginica'])
```

```r
par(new=TRUE)
micro.roc <- roc(y.labels, y.probs, auc.polygon=TRUE, max.auc.polygon=TRUE, print.auc=TRUE, sh
micro.smoothroc <- smooth(micro.roc, method = "density")
plot(micro.smoothroc, col = 'black', lty = 'dotdash', xaxt='n', xlab="", ylab = "")

macro.sensitivity <- (setosa.smoothroc$sensitivities + versicolor.smoothroc$sensitivities + vi
macro.specificity <- ((setosa.smoothroc$specificities + versicolor.smoothroc$specificities + v
lines(macro.specificity, macro.sensitivity, type='l', xlim = rev(range(macro.specificity)), co
axis(1, at=(5:0) * 0.2, labels=(0:5) * 0.2, pos=c(-0.04,0))

legend(0.4, 0.5, legend = c('setosa','versicolor', 'virginica', 'micro-avg', 'macro-avg'), col

probs.train <- predict(model, x.train, type="raw")
multiclass.roc(y.train, probs.train)

multiclass.roc(y.test, probs)

probs.all <- predict(model, iris[1:4], type="raw")
multiclass.roc(iris$Species, probs.all)
```


True Positive Rate (Sensitivity)
0.0
0.2
0.4
0.6
0.8
1.0
0
0.2
0.4
0.6
0.8
1
False Positive Rate (1 - Specificity)
setosa
versicolor
virginica
micro-avg
macro-avg