

# BCB570 Assignment 7

*Ashish Jain*

## 5-fold cross-validation to design a classifier to separate B and T cells in ALL dataset

In this problem, first, we downloaded the ALL data from Bioconductor package. After that, we process the data and created labels for B and T cells. Then, we used the `cvpred.ksvm` function from the `kernlab` package for 5-fold cross-validation. First, we used the linear classifier for the classification.

```
# source('https://bioconductor.org/biocLite.R')
# biocLite('ALL')
library(kernlab)
library(ALL)
library(ROCR)
source("general.R")
data(ALL)
# Predict type of disease from expression data
y <- substr(ALL$BT, 1, 1)
x <- t(exprs(ALL))

k = 5
ypredscorecv <- cvpred.ksvm(x, y, folds = k, type = "C-svc",
  kernel = "vanilladot", C = 1, scaled = c(), predtype = "decision")
```

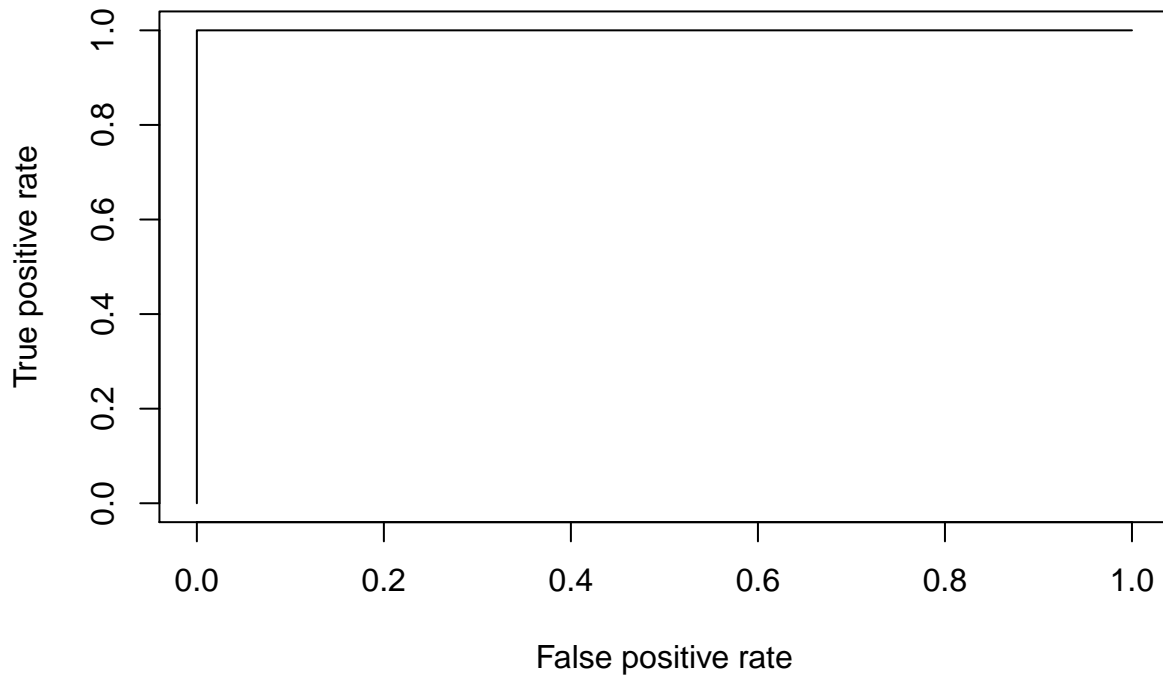
```
## Setting default kernel parameters
## Setting default kernel parameters
## Setting default kernel parameters
## Setting default kernel parameters
## Setting default kernel parameters
```

```
# Check the performance
print(table(ypredscorecv > 0, y))
```

```
##           y
##           B  T
## FALSE 95  0
## TRUE   0 33
```

```
pred <- prediction(ypredscorecv, y)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf, main = "5-fold cross-validation for Linear classifier")
```

## 5-fold cross-validation for Linear classifier



```
# plot(unlist(perf@x.values), unlist(perf@y.values))
```

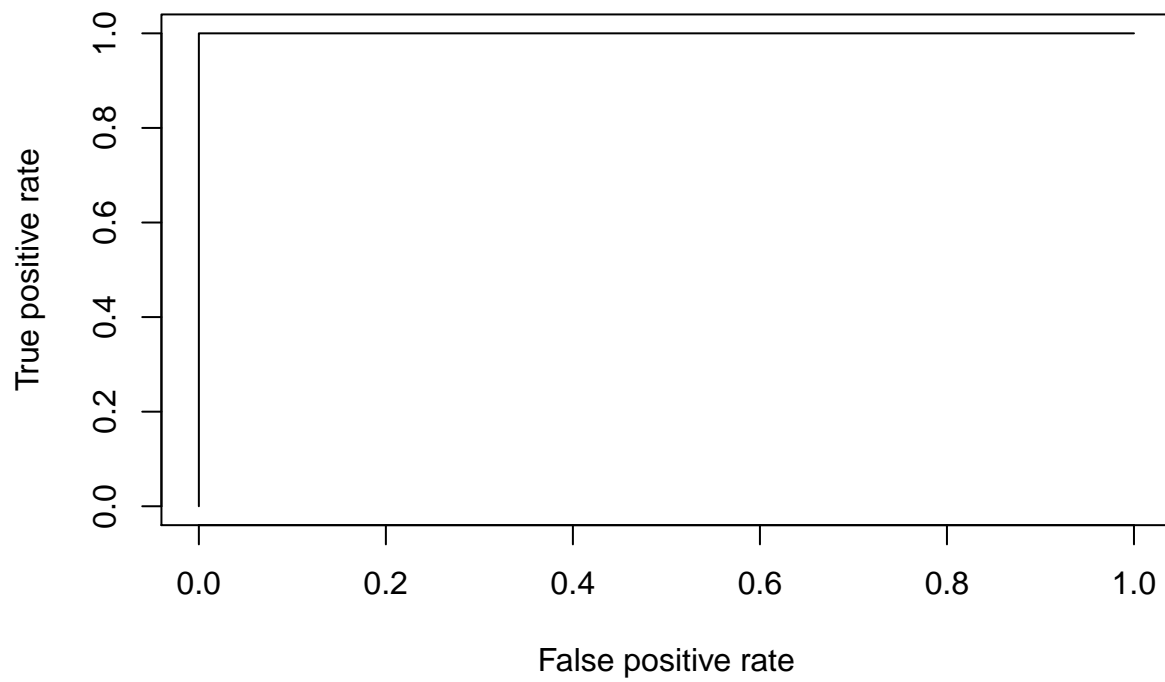
The results show that the linear SVM classifier is able to classify both the B and T cells perfectly (with accuracy equals to 1). The AUROC is also equal to 1. Next, we also tried non-linear classifier using the Radial Basis kernel function.

```
## Non-linear classifier
k = 5
ypredscorecv <- cvpred.ksvm(x, y, folds = k, type = "C-svc",
  kernel = "rbf", C = 1, scaled = c(), predtype = "decision")
# Check the performance
print(table(ypredscorecv > 0, y))
```

```
##           y
##           B  T
## FALSE 95  0
## TRUE   0 33
```

```
pred <- prediction(ypredscorecv, y)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf, main = "5-fold cross-validation for Non-linear classifier")
```

### 5-fold cross-validation for Non-linear classifier



The results show that the non-linear SVM classifier also detects both the B and T cells perfectly (with accuracy equals to 1). The AUROC is also equal to 1.