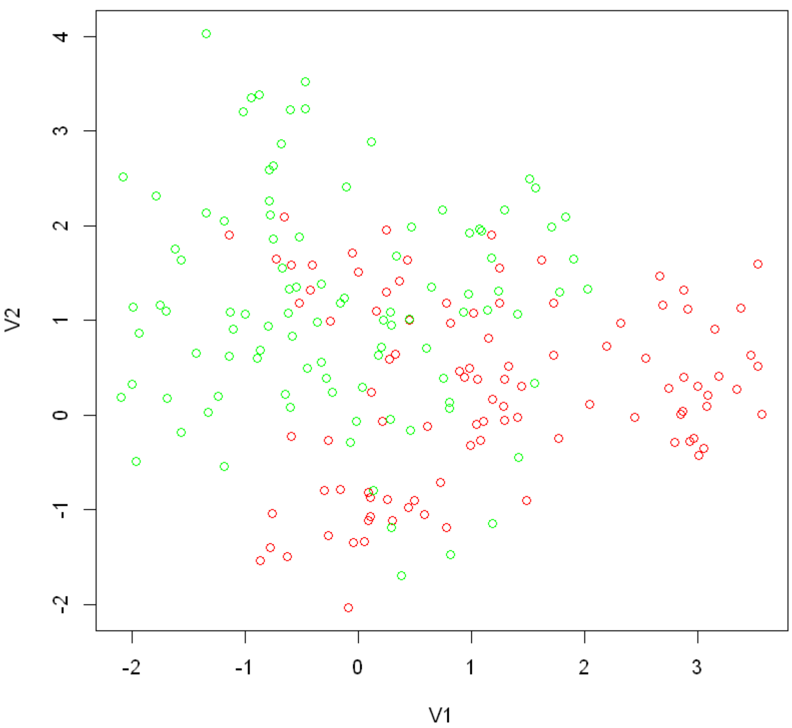
In [1]:

```
setwd("C:/Data/Graduate/Data mining/data")


redpoints = read.table("trainred.txt", sep = "₩t", header = F)
greenpoints = read.table("traingreen.txt", sep = "₩t", header = F)
redtestpoints = read.table("testred.txt", sep = "₩t", header = F)
greentestpoints = read.table("testgreen.txt", sep = "₩t", header = F)

X = rbind(redpoints, greenpoints)
y = c(rep(1, 100), rep(0, 100))
```

In [2]:

```
plot(X, type = "n")

points(redpoints, col = "red")
points(greenpoints, col = "green")
```



In [3]:

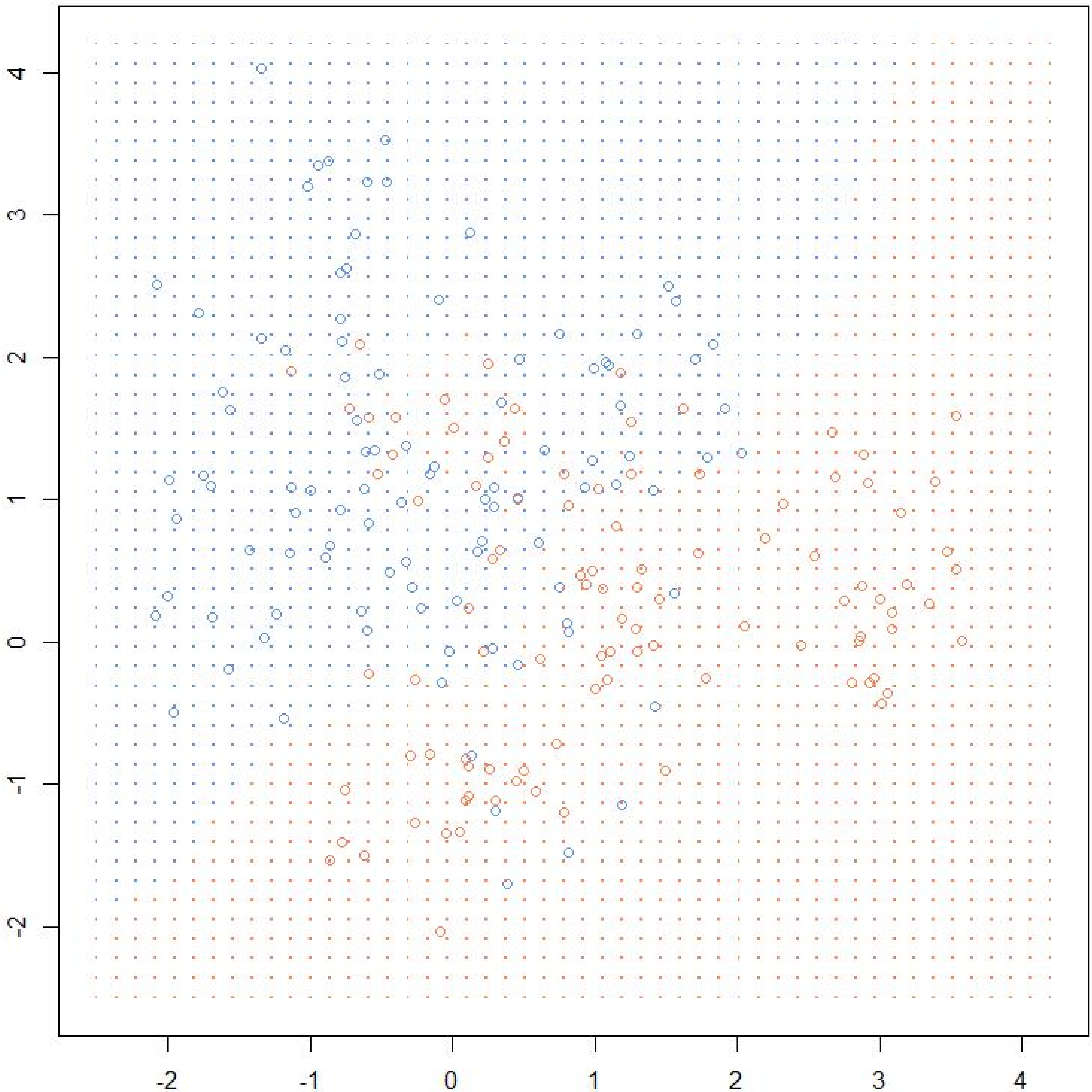
```
# 1.

## Euclidean distance matrix
D = matrix(0, 200, 200)

Dist <- function(Mat1, Mat2, Mat3=NA){
  num_col <- nrow(Mat2)
  if(is.na(Mat3)[1]){
    num_row <- nrow(Mat2)
    Index <- cbind(rep(c(1:num_row), num_col), rep(c(1:num_col), each = num_row))
    list_of_values <- lapply(1:(num_row * num_col), function(n){
      index <- Index[n,]
      sqrt(sum((Mat2[index[1], ] - Mat2[index[2], ])^2))
    })
    Mat1 <- matrix(Reduce(rbind, list_of_values), num_row, num_col)
  }else{
    num_row <- nrow(Mat3)
    Index <- cbind(rep(c(1:num_row), num_col), rep(c(1:num_col), each = num_row))
    list_of_values <- lapply(1:(num_row * num_col), function(n) {
      index <- Index[n,]
      sqrt(sum((Mat3[index[1], ] - Mat2[index[2], ])^2))
    })
    Mat1 <- matrix(Reduce(rbind, list_of_values), num_row, num_col)
  }
  return(Mat1)
}


D <- Dist(D, X)
```

In [ ]:

```
## test data set for depicting grids in the plot

D0 <- matrix(0,2500,200)
X0 <- matrix(0, 2500, 2)
x0 <- seq(from = -2.5, to = 4.2, length = 50)

for(i in 1:length(x0)){
  X0[((i-1)*50+1):(i*50), 1] <- x0[i]
  X0[((i-1)*50+1):(i*50), 2] <- x0
}

A <- Dist(D0, X, X0)
```

In [ ]:

```
## K-nearest neighbor algorithm with k = 15
k=15

g.hat15 = rep(0, 200)
for (i in 1:200){
  g.hat15[i] = (mean(y[order(D[i, ])[1:k]]) > 0.5)
}

training.error = 1-sum(g.hat15 == y)/200
training.error
g.hat15


## K-nearest neighbor algorithm with k = 15 for the test data
g0.p =rep(0,2500)
for (i in 1:2500) g0.p[i]=mean(y[order(A[i,])[1:k]])
g0.hat <- g0.p > 0.5
g0.p

## draw a plot with classified test data
X0
prob_mat <- matrix(g0.p, length(x0), length(x0), byrow = T)

par(mar=rep(2,4))

plot(x0, x0, type = 'n')

points(X, col=ifelse(y == 1, "coral", "cornflowerblue"))
gd <- expand.grid(x=x0, y=x0)
points(gd, pch=".", cex=2, col=ifelse(prob_mat > 0.5, "coral", "cornflowerblue"))
box()
```



4
3
2
1
0
-1
-2
-2
-1
0
1
2
3
4

2. $\min_k \|t_k - \hat{y}\|$, $\quad t_k = e_k = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^K$: Large K.

Define $\hat{y} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_K \end{pmatrix}$. $\sum_{i=1}^{K} \hat{y}_i = 1$, suppose $\hat{y}_i \geq 0 \quad \forall\, i = 1 \cdots K$.

Then $\|t_k - \hat{y}\| = \sqrt{(t_k-\hat{y})'(t_k-\hat{y})} = \sqrt{\hat{y}'\hat{y} - 2\hat{y}_k + 1}$ $\quad$ (trivially, $\hat{y}'\hat{y} - 2\hat{y}_k + 1 \geq 0$)

Since $\sqrt{\ }$ is a strictly increase function at non-negative domain,

$$\min_k \|t_k - \hat{y}\| = \min_k (\hat{y}'\hat{y} - 2\hat{y}_k + 1) = \min_k \{-\hat{y}_k\} = \max_k \{\hat{y}_k\}$$

3. show $\hat{f}(x_0) = \sum_{i=1}^{N} l_i(x_0; \mathcal{X}) y_i$, $\quad y_i = f(x_i) + \varepsilon_i$

1) Linear Regression

Let, $\mathcal{X} \in \mathbb{R}^{N \times p}$ design matrix. $x_i \in \mathbb{R}^p$: fixed value vector. $i = 1 \cdots N$

$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^N$: Response Vector. $\beta \in \mathbb{R}^p$: Regression coefficient.

$f(x_i) = x_i\beta$ under Model assumption

$$\underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|Y - \mathcal{X}\beta\|_2^2 = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \left(\|\mathcal{X}(\hat{\beta}-\beta)\|_2^2 + 2(Y-\mathcal{X}\hat{\beta})'(\mathcal{X}(\hat{\beta}-\beta)) + \|Y-\mathcal{X}\hat{\beta}\|_2^2 \right) \quad \text{where } \hat{\beta} = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'Y$$

$$= \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \left(\|\mathcal{X}(\hat{\beta}-\beta)\|_2^2 \right) \quad \because (Y-\mathcal{X}\hat{\beta})'(\mathcal{X}(\hat{\beta}-\beta)) = Y'(I - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}')\mathcal{X}(\hat{\beta}-\beta) = 0$$

$$= \hat{\beta} = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'Y$$

$$\therefore \hat{f}(x_0) = x_0'\hat{\beta} = \underset{1\times p}{x_0'}\underset{p \times p}{(\mathcal{X}'\mathcal{X})^{-1}}\mathcal{X}'Y = x_0'(\mathcal{X}'\mathcal{X})^{-1}\sum_{i=1}^{N} x_i y_i = \sum_{i=1}^{N} x_0'(\mathcal{X}'\mathcal{X})^{-1}x_i y_i$$

$$\mathcal{X} = \begin{pmatrix} x_1' \\ \vdots \\ x_N' \end{pmatrix} \quad \mathcal{X}' = \begin{pmatrix} \boxed{x_1} \cdots x_N \end{pmatrix} \qquad \therefore l_i(x_0; \mathcal{X}) = x_0'(\mathcal{X}'\mathcal{X})^{-1}x_i, \quad i = 1, \cdots, N$$

2) KNN

$$\hat{f}(x_0) = \frac{1}{K}\sum_{x_i \in N_k(x_0)}^{N} y_i = \frac{1}{K}\sum_{i=1}^{N} I_{(x_i \in N_k(x_0))} y_i$$

$$= \sum_{i=1}^{N} \underbrace{\frac{1}{K} I(x_i \in N_k(x_0); \mathcal{X})}_{l_i(x_0; \mathcal{X})} y_i$$

4. Let, $X \in \mathbb{R}^p$, $Y \in \mathbb{R}$ random variable.

$$EPE(f) = E[(Y-f(X))^2] . \quad \text{show} \quad \underset{f}{\operatorname{argmin}} \{EPE(f)\} = E[Y|X]$$

i) Lemma. $Cov(Y-E[Y|X], f(X)) = 0$, $\forall f(X)$ (임의의 X에 대한 함수)

pf) Let, $Z = Y - E[Y|X]$. Then $E[Z|X] = E[Y|X] - E[E[Y|X]|X] = E[Y|X] - E[Y|X] = 0$

$\therefore E[Z] = E[E[Z|X]] = E[0] = 0$ $\because$ double expected value theorem

$$Cov(Z, f(X)) = E[Z \cdot f(X)] - \overset{0}{\cancel{E[Z]}} E[f(X)] = E[E[Z f(X)|X]] = E[f(X) \overset{0}{\cancel{E[Z|X]}}] = E[0] = 0 \quad \square$$

ii) Claim: $E[(Y-f(X))^2] \geq E[(Y-E[Y|X])^2]$, $\forall f(X)$

$$Y - f(X) = (Y - E[Y|X]) \oplus (E[Y|X] - f(X)) \qquad \oplus : \text{direct sum.}$$

Let, $u(X) = E[Y|X] - f(X)$

$$(Y-f(X))^2 = (Y-E[Y|X] \oplus u(X))^2 = (Y-E[Y|X])^2 + 2(Y-E[Y|X])u(X) + (u(X))^2$$

Hence, $E[(Y-f(X))^2] = E[(Y-E[Y|X])^2] + 2E[\overbrace{(Y-E[Y|X])\, u(X)}^{Cov(Y-E[Y|X],\, u(X))}] + E[u(X)^2]$

$= E[(Y-E[Y|X])^2] + E[u(X)^2]$ $\because$ Lemma

$\geq E[(Y-E[Y|X])^2]$, $\forall f(X)$, which determines $u(X)$ $\therefore$ claim: True.

Hence. $E[(Y-E[Y|X])^2]$ is infimum of EPE on the space of f.

$$\therefore \underset{f}{\operatorname{argmin}}\, E[(Y-f(X))^2] = E[Y|X].$$