EECS Bayes Model in Machine Learning

JINGYING ZHOU

JZ2673

Problem 1 If one swithes door, his/her chance of winning ~~ex~~ is:

Should switch!

~~P(A2)~~ Set Event A: Start (off) with the right box.
Event B: swith to the right box and win prize.

$$P(B) = P(A)\,P(B|A) + P(B|\bar{A})\cdot P(\bar{A})$$
$$= \frac{1}{3}\times 0 + \frac{2}{3}\times 1 = \frac{2}{3}$$

Event C: Do not swith box and win the prize.

$$P(C) = P(C|A)\cdot P(A) + P(C|\bar{A})\cdot P(\bar{A}) = \frac{1}{3}\times 1 + \frac{2}{3}\times 0 = \frac{1}{3}$$

∴ The probability of ~~start~~ winning prize under swithing strategy is $\frac{2}{3}$. So we should advise him to switch.

Problem 2 The ~~mo~~ multinomial likelihood gives

~~$f(x_1 \ldots x_k; n, p_1 \ldots p_k) = Pr(X_1 = x_1 \ldots X_k = x_k) =$~~

$$P(\vec{X}|\pi_1 \ldots \pi_k) = \prod_{i=1}^{N} P(X_i = x_i|\pi_1 \ldots \pi_k) = \prod_{i=1}^{N}\prod_{j=1}^{k} \pi_j^{I(x_i=j)}$$

$$= \prod_{j=1}^{k} \pi_j^{n_j}$$

where $n_j = \sum_{i=1}^{N} I(x_i = j)$

Then $P(\vec{X}|\pi_1 \ldots \pi_k) = e^{\sum_{i=1}^{k} n_j \log(\pi_j)}$

In order to find a conjugate prior, we hope to find the form of $e^{\sum_{i=1}^{k}[\log \pi_i (\text{constant})]}$

And we know dirichlet distribution's pdf is

$$\frac{1}{B(\alpha)}\prod_{i=1}^{k} x_i^{\alpha_i - 1} = \frac{1}{B(\alpha)} e^{\sum_{i=1}^{k}(\alpha_i - 1)\log x_i}$$ substitute x as π.

$$P(\vec{\pi}) = \frac{1}{B(\alpha)} e^{\sum_{i=1}^{k}(\alpha_i - 1)\log \pi_i}$$

Then the posterior is in the form of

$$P(\pi \mid x_1 \ldots x_N) \propto P(x_1 \ldots x_N \mid \pi)\, P(\pi) \propto e^{\sum_{i=1}^{k} \log(x_i)(\alpha_i + N_i - 1)}$$

$$\propto \text{Dirichlet}(\alpha_1+n_1, \alpha_2+n_2, \ldots \alpha_k+n_k$$

~~where~~ $\cancel{n_j = \sum_{i=1}^{N}}$

$\cancel{A_j = \sum^{N}}$

where $n_i = \sum_{j=1}^{N} 1(x_j = i)$

The posterior parameters, for i in $1 \cdots k$, $\alpha_i^{Post} = \alpha_i + n_i$
which incoporates our belief in prior and the actual data.

As we collect more data, the posterior distribution of parameter will agree more with data and less with prior we set.

EECS Bayes Model in Machine Learning
JINGYING ZHOU
JZ2673

Problem 1

Should switch!

REPEATED SCAN!
PLEASE IGNORE!

-

-

-

-

SORRY &
Have a nice one :D

Problem

EECS E6720 HW1
JINGYING ZHOU
UNI: JZ2673

Problem 3 a) ① $P(\mu|\lambda,\vec{x}) = \frac{P(\vec{x}|\mu,\lambda)P(\mu|\lambda)}{P(\vec{x}|\lambda)} \propto P(\vec{x}|\mu,\lambda)P(\mu|\lambda)$

for $P(x_1\cdots x_n|\lambda)$ is not a function of μ,

since $P(x_1\cdots x_n|\lambda) = \int P(\mu, x_1\cdots x_n|\lambda)d\mu$ which integrates μ out

Then $P(\mu|\lambda,x_1\cdots x_n) \propto \prod_{i=1}^{N} \sqrt{\lambda}\, e^{-\frac{\lambda}{2}(y_i-\mu)^2} \sqrt{\frac{\lambda}{a}}\, e^{-\frac{\lambda}{2a}\mu^2}$

$\propto \lambda^{\frac{N+1}{2}} e^{-\frac{\lambda}{2a}[(1+aN)\mu^2 - 2a\sum_{i=1}^{N} y_i\cdot\mu]}$

$\propto \lambda^{\frac{N+1}{2}} e^{\frac{\lambda(1+aN)}{2a}(\mu - \frac{a\sum x_i}{1+aN})^2}$

This is in a normal form with mean: $\frac{a\sum_{i=1}^{N} x_i}{1+aN}$

variance: $\frac{a}{1+aN}\lambda^{-1}$

$\therefore \mu|x_1\cdots x_N,\lambda \sim Normal\left(\frac{aN\bar{x}}{1+aN}, \frac{a}{1+aN}\lambda^{-1}\right)$

② $P(\lambda|x_1\cdots x_N) = \int P(\mu,\lambda|\vec{x})d\mu = \int P(x_1\cdots x_N|\mu,\lambda)P(\mu|\lambda)P(\lambda)\,d\mu$

$= P(\lambda)\int P(x_1\cdots x_N|\lambda,\mu)P(\mu|\lambda)d\mu$

$\propto P(\lambda)\int \lambda^{\frac{N+1}{2}} a^{\frac{1}{2}} e^{-\frac{\lambda}{2a}(\mu^2 + a\sum_{i=1}^{N}(x_i-\mu)^2)}\, d\mu$

$\propto P(\lambda)\lambda^{\frac{N+1}{2}} a^{\frac{1}{2}} \int e^{-\frac{\lambda}{2a}[\mu^2 + a\sum_{i=1}^{N}(x_i-\mu)^2]} d\mu$

Complete Squares: $\propto P(\lambda)\lambda^{\frac{N+1}{2}} a^{\frac{1}{2}} \cdot \underbrace{\int e^{-\frac{1}{2}\left(\frac{\lambda(1+aN)}{a}\right)\left(\mu - \frac{a\sum x_i}{1+aN}\right)^2} d\mu}_{①} \cdot e^{-\frac{\lambda}{2}\left(\frac{1+aN}{a}\right)\left[\frac{a\sum x_i^2}{1+aN} - \frac{a^2(\cdots}{(1+a\cdots}\right.}$

Part ①. i.e. the integral part follows ~~a~~ the form of normal pdf. so the integral is proportional to σ, where $\sigma = \sqrt{\frac{a}{(1+aN)\lambda}}$

We then have:

$$P(\lambda|y) \propto \underbrace{P(\lambda)\cdot\lambda^{\frac{N+1}{2}} a^{\frac{1}{2}} \sqrt{\frac{a}{(1+aN)\lambda}}\cdot 1}_{\text{Part ①}} \cdot \underbrace{e^{-\frac{\lambda}{2}\left[\sum^{N} x_i^2 - \frac{a}{1+aN}(\sum x_i)^2\right]}}_{\text{Part ②}}$$

$$\text{part ①} \propto c^b \lambda^{b-1+\frac{N}{2}} e^{-c\lambda} \cdot a \propto \lambda^{b+\frac{N}{2}-1} e^{-c\lambda}$$

$$\text{Part ②} \propto e^{-\frac{\lambda}{2}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2 - \sum_{i=1}^{N} N\bar{x}^2 + 2\bar{x}\sum_{i=1}^{N} x_i - \frac{aN}{1+aN}(N\bar{y})^2\right]}$$

$$= e^{-\frac{\lambda}{2}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2 + N\bar{x}^2\left(1-\frac{aN}{1+aN}\right)\right]}$$

$$= e^{-\frac{\lambda}{2}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{N\bar{x}^2}{1+aN}\right]}$$

$$P(\lambda|x_1 \ldots x_n) \propto \text{Part①}\cdot\text{Part②} \propto \lambda^{b+\frac{N}{2}-1} \cdot e^{-\lambda\left[c+\frac{1}{2}\left(\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{N}{1+aN}\bar{x}^2\right)\right]}$$

$$\propto \text{Gamma}(\alpha, \beta)$$

where $\alpha = b+\frac{N}{2}$

$$\beta = c + \frac{1}{2}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{N}{1+aN}\bar{x}^2\right] \quad \text{where } \bar{x} = \sum_{i=1}^{N} x_i$$

Prob 3b) $P(x^*|x_1 \ldots x_n) = \int_0^\infty \int_{-\infty}^\infty P(x^*|\mu,\lambda) P(\mu,\lambda|x_1 \ldots x_N)\, d\mu \cdot d\lambda$

$$= \int_0^\infty \int_{-\infty}^\infty \text{Normal}(\mu, \lambda^{-1}) \cdot \text{Normal}(g, h\lambda^{-1}) \cdot \text{Gamma}(\alpha, \beta)\, d\mu \cdot d\lambda$$

where: g, h, α, β are all functions of $x_1 \ldots x_n$.

$$g = \frac{aN}{1+aN}\bar{x}$$

$$h = \frac{a}{1+aN}$$

Then $P(x^*|x_1 \ldots x_n) = \int_0^\infty \int_{-\infty}^\infty \sqrt{\frac{\lambda}{2\pi}} e^{-\frac{\lambda}{2}(x^*-\mu)^2} \sqrt{\frac{\lambda}{2\pi h}} e^{-\frac{\lambda}{2h}(\mu-g)^2} \cdot \text{Gamma}(\alpha\,\beta)\, d\mu\, d\lambda$

$$\propto \int_0^\infty \frac{\lambda}{\sqrt{h}} \text{Gamma}(\alpha, \beta) \int_{-\infty}^\infty e^{-\frac{\lambda}{2}\left[x^{*2}+\mu^2-2x^*\mu+\frac{1}{h}(\mu-g)^2\right]}$$

$$= \int_0^{\infty} \frac{\lambda}{\sqrt{h}} \text{Gamma}(\alpha,\beta) \int_{-\infty}^{\infty} e^{-\frac{\lambda}{2}\left\{\left[(1+\frac{1}{h})(\mu - \cancel{g}\frac{xh+g}{n+1})^2\right] - \frac{(xn+g)^2}{h(1+h)} + x^2 + g^2/h\right\}} \, d\mu \, d\lambda$$

$$\propto \int_0^{\infty} \frac{\lambda}{\sqrt{h}} \text{Gamna}(\alpha,\beta) \sqrt{\frac{h}{(1+h)\lambda}} * e^{-\frac{\lambda}{2}\left[\frac{-(xh+g)^2 + x^2h + x^2h + g^2h^2 + g^2h}{h(h+1)}\right]} \, d\lambda$$

$$= \int_0^{\infty} \lambda^{\frac{1}{2}} \text{Gamma}(\alpha,\beta) \frac{1}{\sqrt{1+h}} e^{-\frac{\lambda}{2}\left[\frac{(x-g)^2}{h+1}\right]} \, d\lambda$$

$$= \left[\int_0^{\infty} \lambda^{\alpha+\frac{1}{2}-1} e^{-\lambda(\beta + \frac{(x-g)^2}{2(h+1)})} \, d\lambda\right] \frac{\beta^{\alpha}}{\sqrt{1+h}\,\Gamma(\alpha)}$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{1}{\sqrt{1+h}} \cdot \frac{\Gamma(\alpha+\frac{1}{2})}{(\beta + \frac{(x-g)^2}{2(h+1)})^{\alpha+\frac{1}{2}}}$$

$$\propto \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \left(\beta(1+h)\right)^{-\frac{1}{2}} \left(1 + \frac{(x-g)^2}{2\beta(h+1)}\right)^{-\alpha-\frac{1}{2}} \cdot \frac{1}{\sqrt{2\pi\alpha}} \cdot \sqrt{\alpha}$$

$$= \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \frac{1}{\sqrt{2\pi\alpha} \cdot \sqrt{\frac{\beta(1+h)}{\alpha}}} \left(1 + \frac{1}{2\alpha} \cdot \frac{(x-g)^2}{\frac{2\beta(h+1)}{2}}\right)^{-\alpha-\frac{1}{2}} \sim$$

Which is a non-standardized t-distribution with $\nu = 2\alpha$

$$\mu = g$$

$$\sigma^2 = \frac{\beta(1+h)}{2}$$

Bayes ML HW1

Jingying Zhou jz2673

September 30, 2016

Problem 4

a) I used the following code to implement the algorithm.

```
### Question a)
rm(list=ls())
Xtrain <- read.csv("/Users/Bianbian/Downloads/hw1_data_csv/Xtrain.csv", header = FALS
E)
Xtest <- read.csv("/Users/Bianbian/Downloads/hw1_data_csv/Xtest.csv",header = FALSE)
Ytrain <- read.csv("/Users/Bianbian/Downloads/hw1_data_csv/Ytrain.csv", header = FALS
E)
Ytest <- read.csv("/Users/Bianbian/Downloads/hw1_data_csv/Ytest.csv", header = FALSE)
# Posterior Marginal of New Y
pi.post <- (1+sum(Ytrain))/(dim(Ytrain)[1]+1+1)
# Group by result
x0train <- Xtrain[Ytrain==0,]
x1train <- Xtrain[Ytrain==1,]
# Calculate Grouped Parameters
param<-function (df){
  N <- dim(df)[1]
  d <- dim(df)[2]

  colvar <- apply(df,2,var)*(N-1)/N
  colmean<- apply(df,2,mean)

  alpha <- rep(1+N/2,d)
  beta <-1+1/2*(N*colvar +N/(1+N)*colmean^2)

  g <- colmean*N/(1+N)
  h <- rep(1/(1+N),d)

  nu = rep(2+N,d)
  miu = g
  sig2<-(beta*(1+h))/alpha
  return(list(nu,miu,sig2))
}

# Assign param values
par0<-param(x0train)
```

```
par1<-param(x1train)
# Pdf of T
tdist<- function(df,par){
  nu <- par[[1]]
  miu <-par[[2]]
  sig2<-par[[3]]
  p = lgamma((nu+1)/2)-lgamma(nu/2) -1/2*(log(nu)+log(sig2))-(nu+1)/2*log(1+1/nu*(df-
miu)^2/sig2)
  return(prod(exp(p)))
}
# Predict if 1 or 0 (9 or 4)
pred<- function(df){
  return(tdist(df,par1)*(pi.post)/(tdist(df,par0)*(1-pi.post)+pi.post*tdist(df,par1))
)
}
```

b) Predicting labels, and print the two way table

```
Ypred <-rep(0,dim(Xtest)[1])
Ypred[apply(Xtest,1,pred)>0.5]=1
# Prediction Accuracy:
ac<- sum(Ypred==Ytest)/dim(Xtest)[1]
print(paste("The prediction accuracy is ",ac))
```

```
## [1] "The prediction accuracy is  0.932697137117027"
```

```
###Question b) Calculate the two way table
# sum(Ypred!=Ytest&Ytest==1)
# sum(Ypred==Ytest&Ytest==1)
# sum(Ypred!=Ytest&Ytest==0)
# sum(Ypred==Ytest&Ytest==0)
tbl <-table(t(Ypred),t(Ytest))
tbl
```

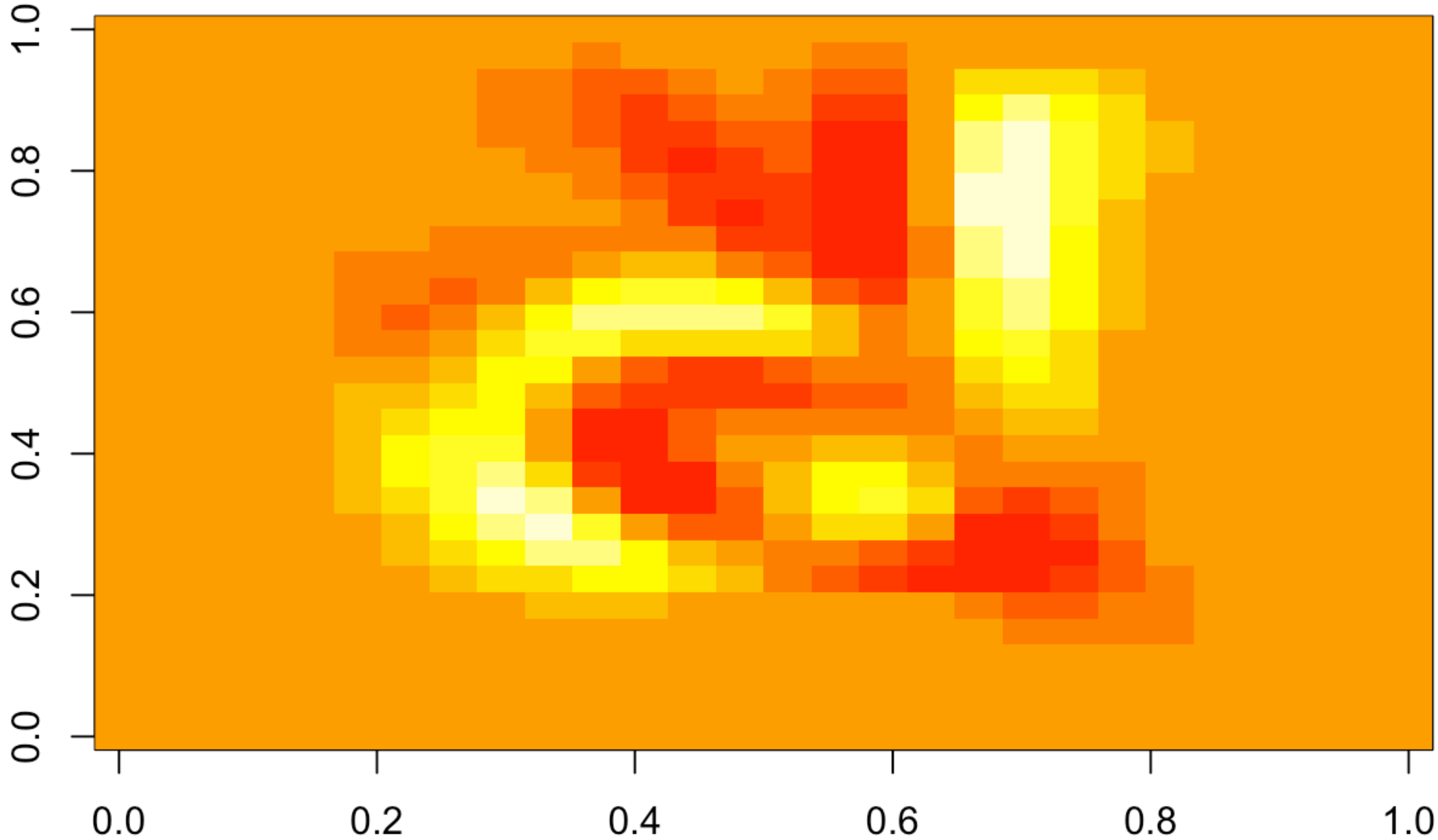
```
##
##       0   1
##   0 930  82
##   1  52 927
```

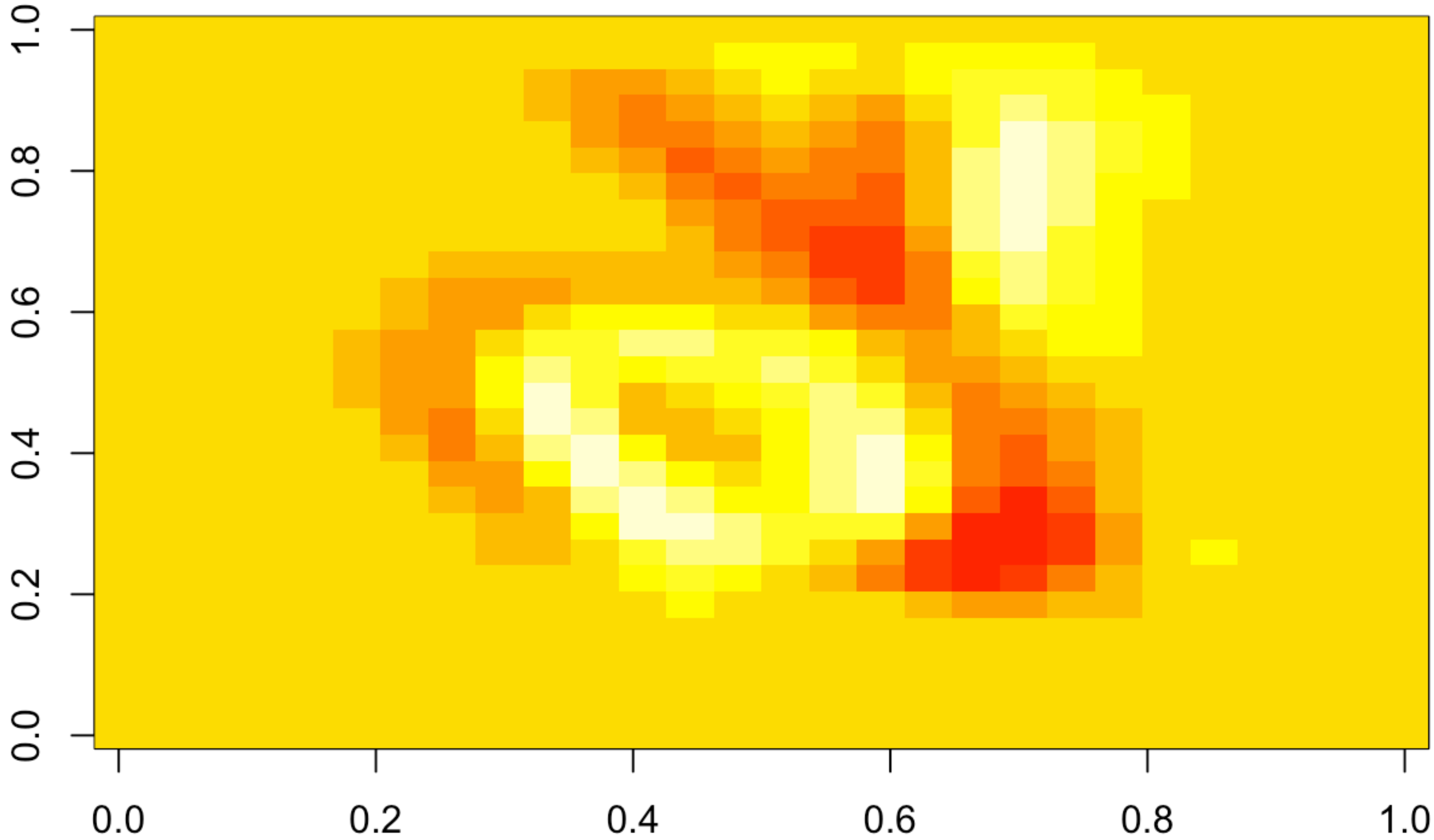
c) I randomly sampled 3 mis-constructed digit:

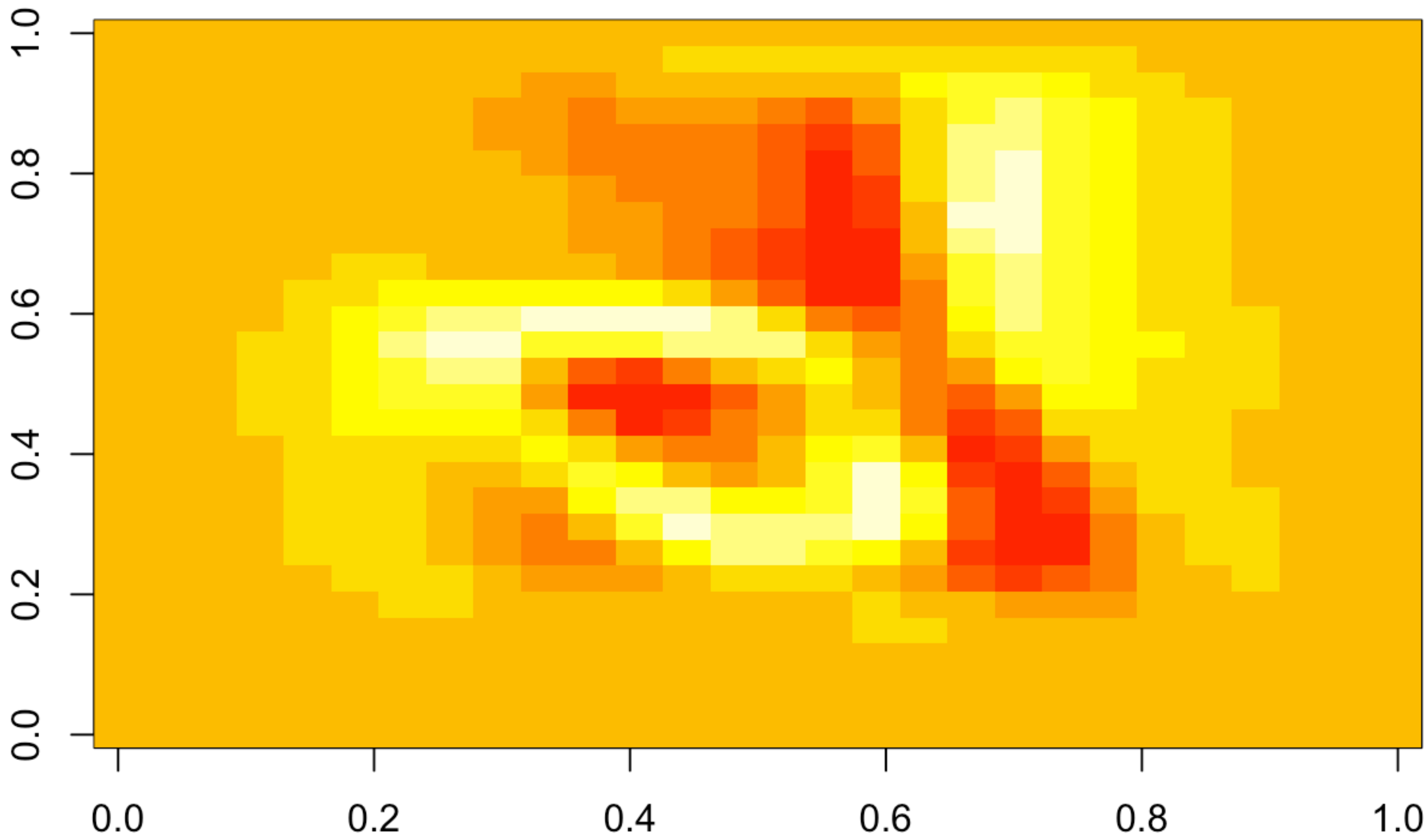
```r
###Question c) Plotting Missclassfied
Q <- read.csv("/Users/Bianbian/Downloads/hw1_data_csv/Q.csv",header=FALSE)
matq <- data.matrix(Q)
matx <- data.matrix(Xtest)
image <- t(matq %*% t(matx))
image.wrong = image[Ypred!=Ytest,]

mat.wrong1 <- matrix(image.wrong[1,],28,28,byrow=FALSE)

for (i in sample(1:134,3)){
  image(matrix(image.wrong[i,],28,28))
}
```


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0



d) The most ambiguous ones are:

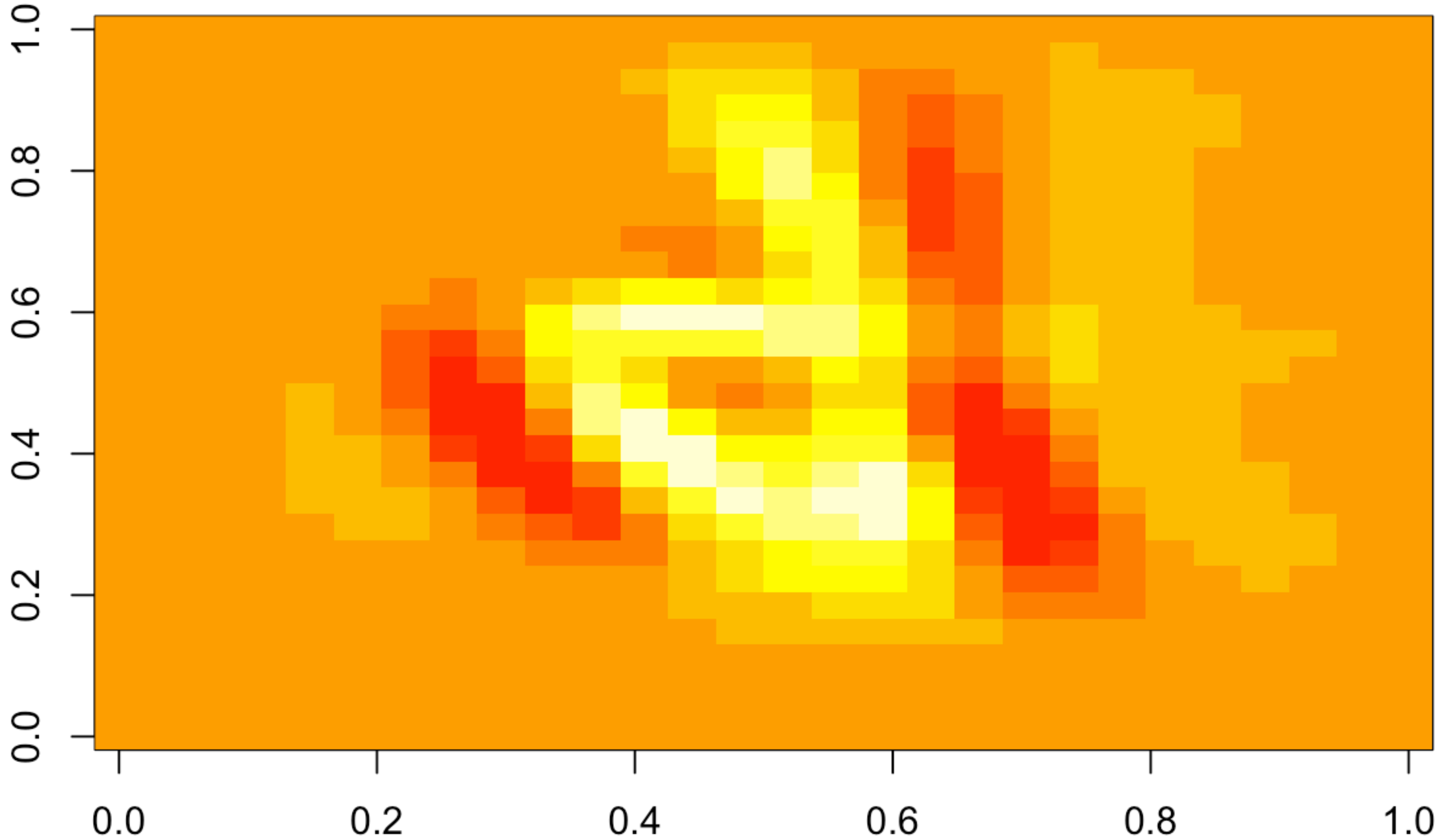
```
### Question d) Plotting Most Ambiguous
# p = apply(Xtest,1,pred)
# ambi<-matx[rank(abs(p-0.5)) <4,]
#
# for (i in 1:3){
#   image(matrix(image.wrong[i,],28,28))
# }
#
#
p = apply(Xtest,1,pred)
ambi<-image[rank(abs(p-0.5)) <4,]
```

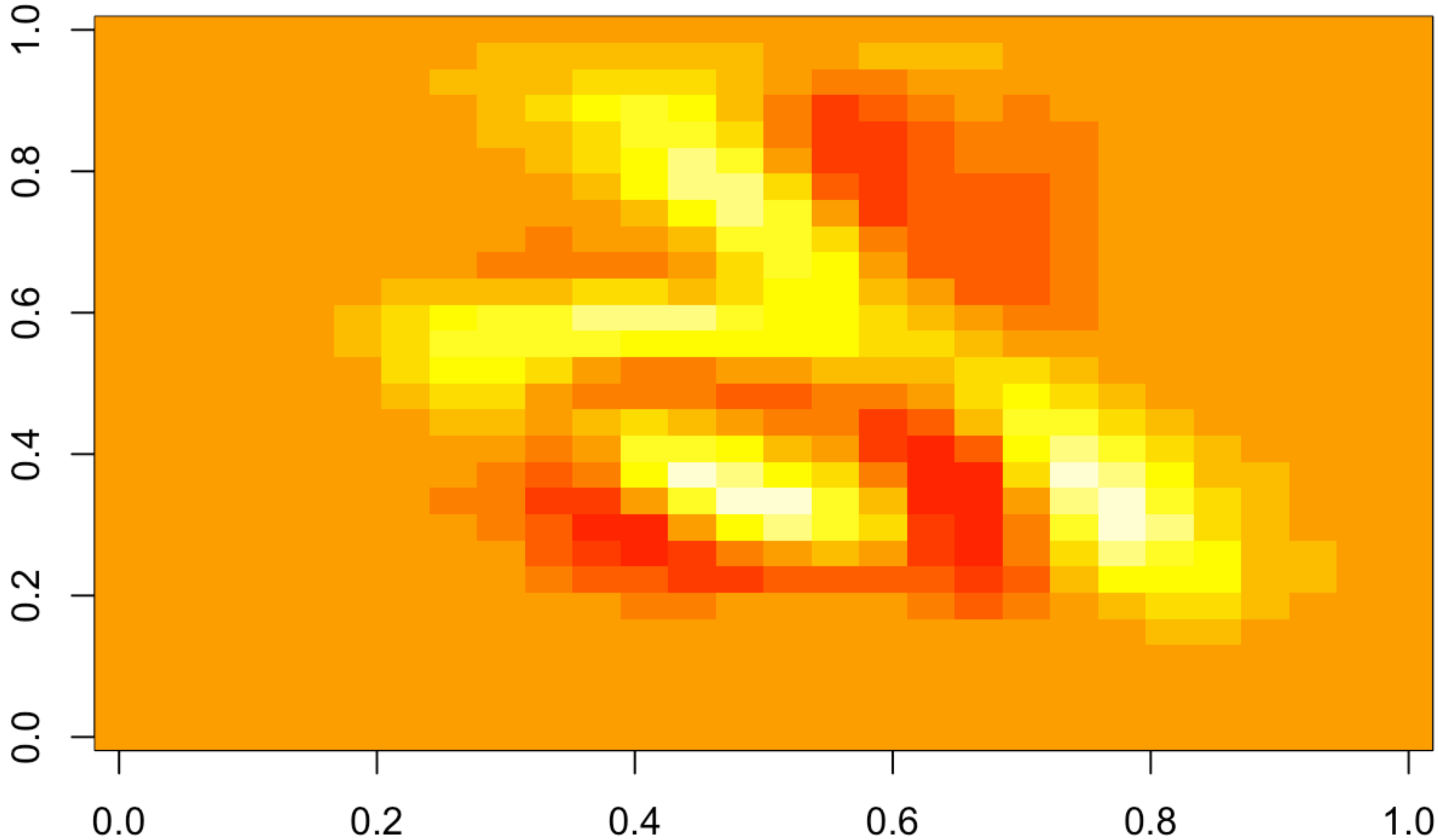
The most ambiguous 3 are:

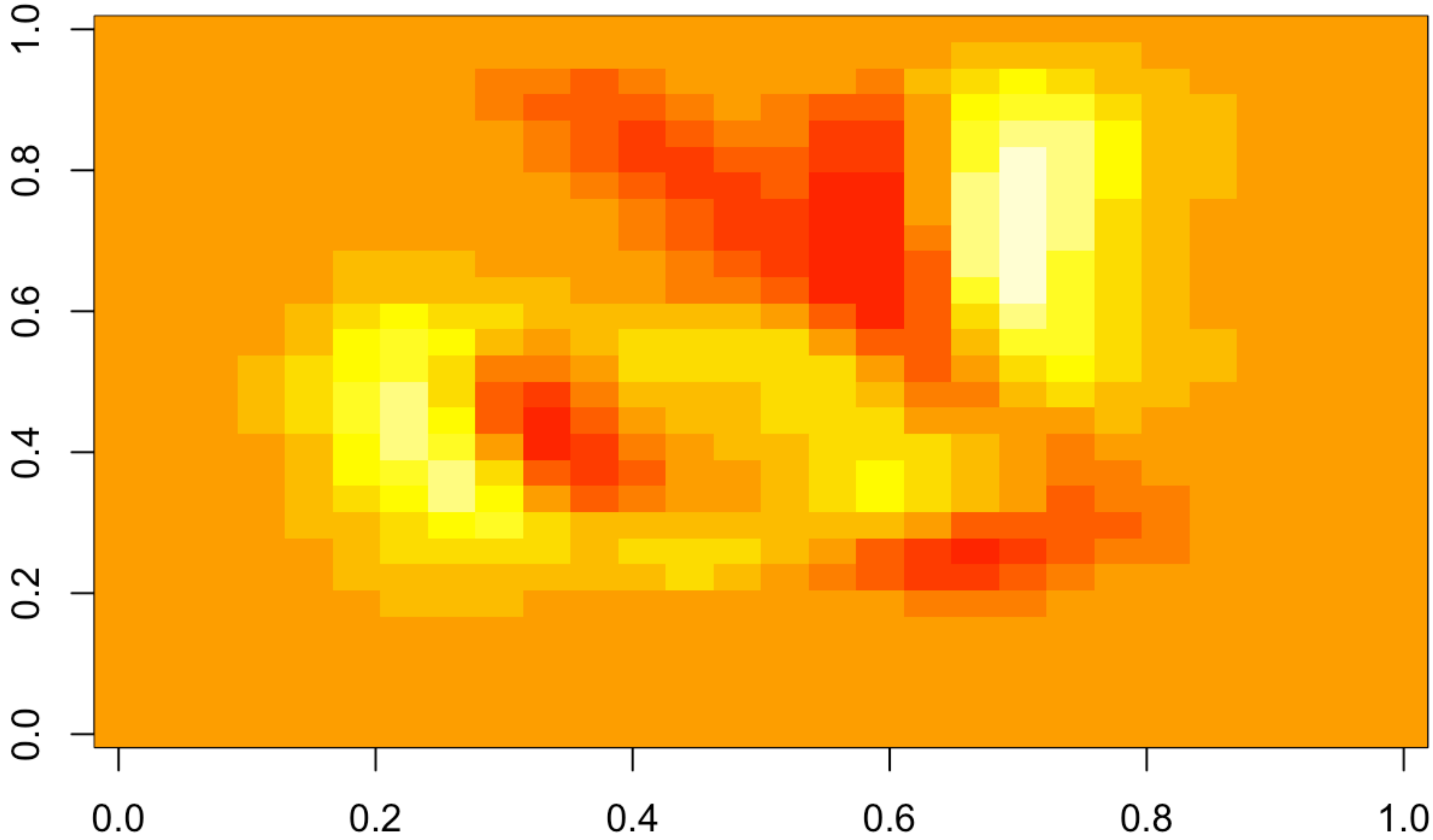
```
print(which(rank(abs(p-0.5)) <4))
```

```
## [1]  358  823 1131
```

```
for (i in 1:3){
  image(matrix(ambi[i,],28,28))
}
```


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0