Non-linear dimensionality reduction

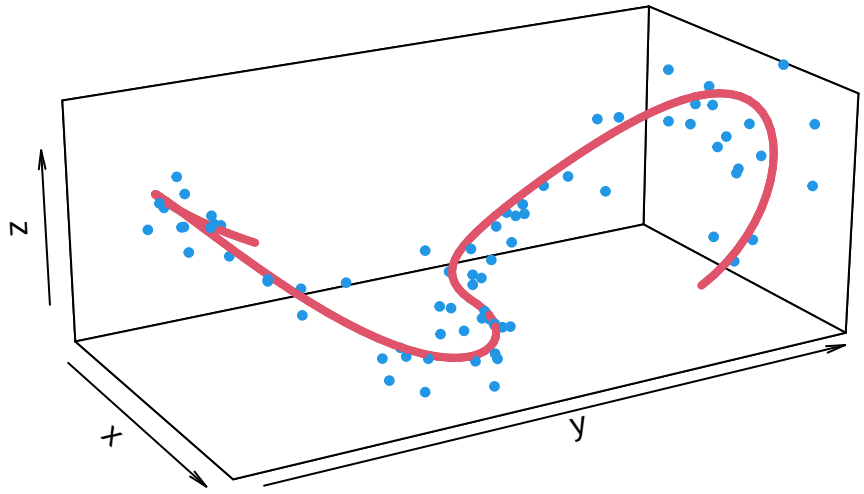
Principal curves, local MDS, Isomap and t-SNE

Caballero Vergés Biel, Menzenbach Svenja and Reyes Illescas Kleber Enrique

2023-10-16

PART A. Principal Curves

1.



Questions

a.

```
df <- seq(2,8, by=1)

#project_to_curve
```

b.

```
# Give a graphical representation of the principal curve output for the optimal df and comment on the o
```

c.

- Before fitting the principal curve with df=50 and based only on the leave-one-out cross-validation error values, what value for df do you think that is better, the previous optimal one or df=50?
- Fit now the principal curve with df=50 and plot the fitted curve in the 3D scatterplot of the original points. Now, what value of df do you prefer?

```
# df = 50
```

- The overfitting with df=50 is clear. Nevertheless leave-one-out cross-validation has not been able to detect this fact. Why do you think that df=50 is given a so good value of leave-one-out cross-validation error?

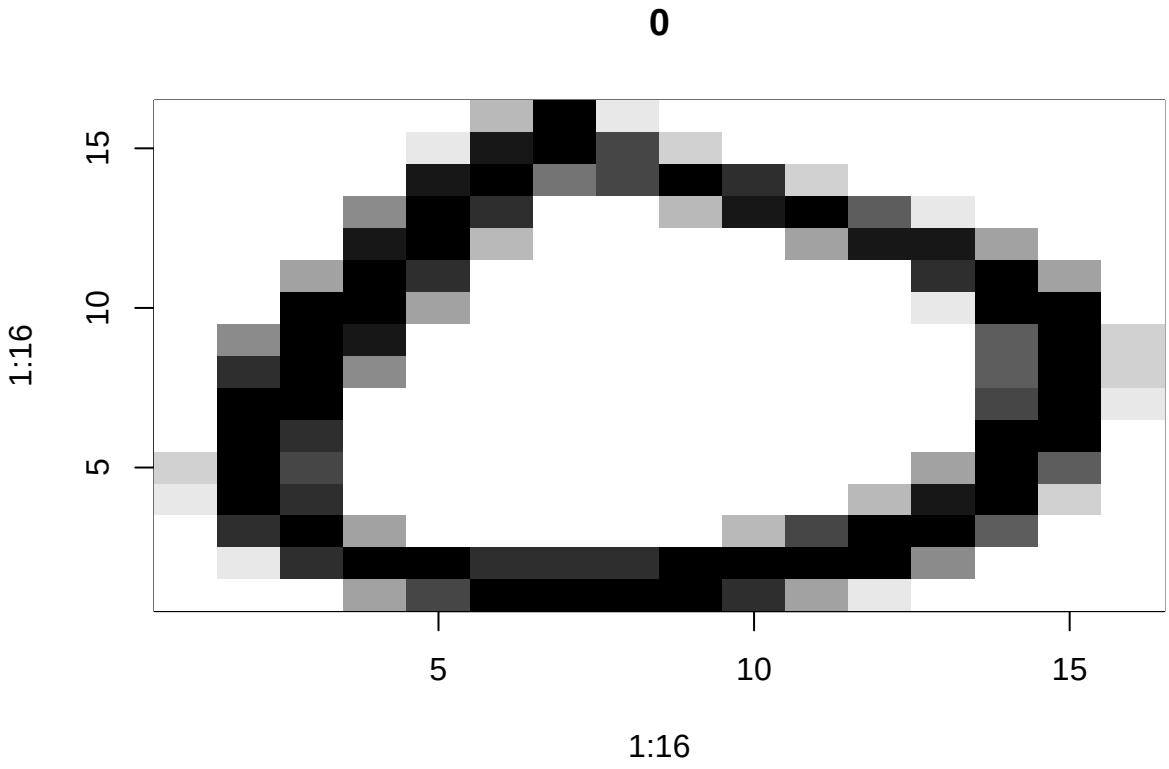
PART B. Local MDS, ISOMAP and t-SNE

```
# ploting 1 digit
plot.zip <- function(x,use.first=FALSE,...){
  x<-as.numeric(x)
  if (use.first){
    x.mat <- matrix(x,16,16)
  }else{
    x.mat <- matrix(x[-1],16,16)
  }
  image(1:16,1:16,x.mat[,16:1],
        col=gray(seq(1,0,l=12)),...)
  invisible(
    if (!use.first){
      title(x[1])
    }else{
    }
  )
  #col=gray(seq(1,0,l=2)))
}
```

2.

```
zip.train <- read.table("zip.train")
zip.train.0 <- zip.train[zip.train[1] == 0,]


plot.zip(zip.train.0[1,]) # plotting the first zero
```



3.

4.

5.