Tutorial: PCA

Benoit Liquet[*1]

[1]Macquarie University

*benoit.liquet-weiland@mq.edu.au

Contents

1 Prediction using PCA

In this question we will use the dataset `Breast_cancer.RData` which contains quantitative information from digitized images of a diagnostic test (fine needle aspirate (FNA) test on breast mass) for the diagnosis of breast cancer. The variables describe characteristics of the cell nuclei present in the image. The mean, standard error and "worst' " or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

```
load("Breast_cancer.RData")
```

(a) Run a PCA model on the variables `radius _mean`, `texture_mean`,`perimeter_mean`, `area_mean`, `smoothness_mean`, `compactness_mean`,`concavity_mean`, `concave.points_mean`, `symmetry_mean`, `fractal_dimension_mean`. We will use the function `prcomp()`.

The prcomp() function will include in its results

- sdev - the standard deviations of the principal components

- rotation - the matrix of variable **loadings** for the components

- x - the scaled matrix of data times the factor loadings (known as scores)

(b) We now can see how much variance is explained by the components. We will use some functions of the package `factoextra` to plot some of the results. The plot representing the variance explained by the components is called `Scree Plot`. Provide the scree plot for this analysis. Comment your results.

(c) Provide the circle-correlation plot for the 3 first components. Comments your results.

(d) Use the eigen value decomposition of the appropriate matrix to get the same result as the `prcomp` function.

(e) Use the singular value decomposition of the appropriate matrix to get the same result as the `prcomp` function.

(f) Project the samples on the two first components. Annotate the sample using the `diagnosis` variables ("Malignant", "Benin").

(g) Use the tree first principal components as predictors in a logistic model for the variable diagnosis. Use the `caret` package and a K-fold cross validation ($K = 10$) approach for providing the sensitivity, specificity and AUC values.

2 Data compression using SVD

In this question we will use the SVD to compress an image. The following code enables us to read a jpeg image file into an array with 3 channels (Red-Green-Blue, RGB).

```
if (!"jpeg" %in% installed.packages()) install.packages("jpeg")
# Read image file into an array with three channels (Red-Green-Blue, RGB)
mates <- jpeg::readJPEG("pool_color.jpg")
```

We then get 3 channels:

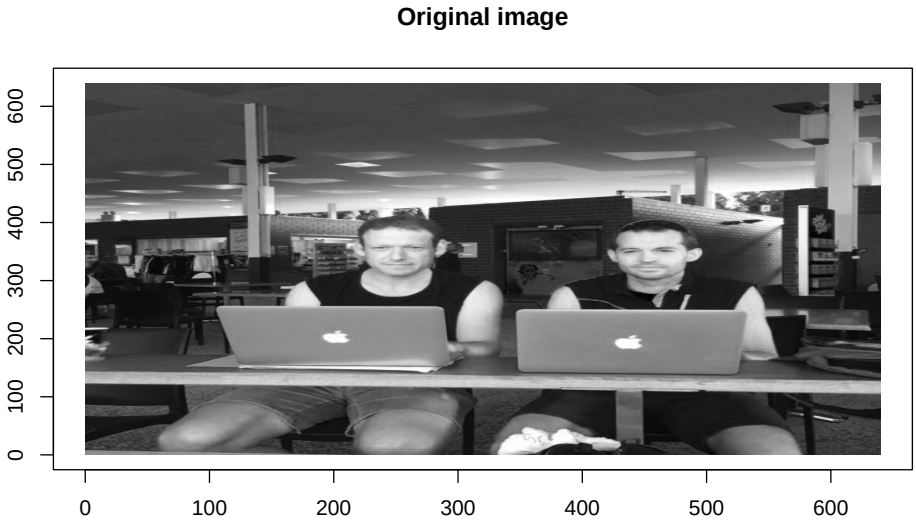
```
r <- mates[, , 1] ; g <- mates[, , 2] ; b <- mates[, , 3]
```

This function enables to plot an image from an array

```
plot.image <- function(pic, main = "") {
  h <- dim(pic)[1] ; w <- dim(pic)[2]
  plot(x = c(0, h), y = c(0, w), type = "n", xlab = "", ylab = "", main = main)
  rasterImage(pic, 0, 0, h, w)
}
```

For simplicity we will only use the red channel information. This will correspond to a black and white image. An image is a matrix of pixels with values that represent the intensity of the the pixel, where $0 = white$ and $1 = black$.

```
plot.image(r, "Original image")
```



This image is a (640×640) pixels

```
dim(r)
[1] 640 640
```

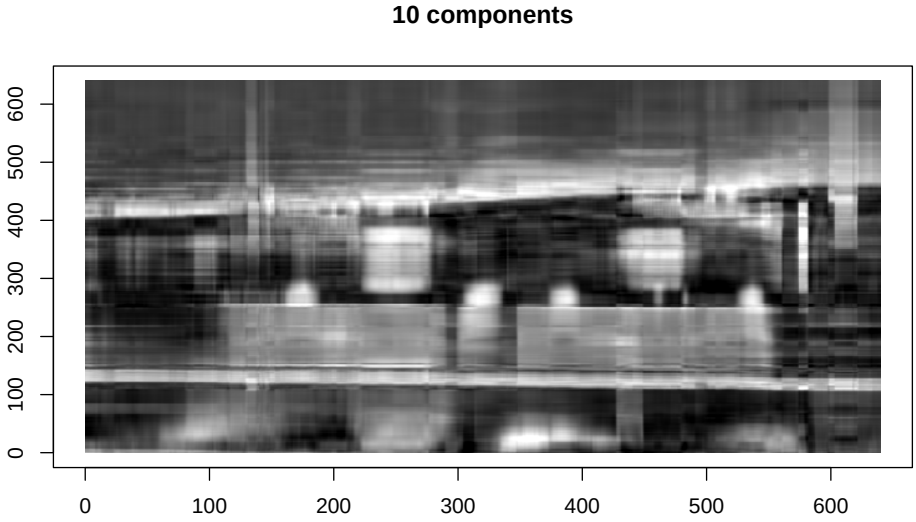
Using the SVD properties (see lecture) we can try to compress the image by using only 10 components.

```
nb.comp <- 10
r.svd <- svd(r)
d <- r.svd$d[1:nb.comp]
v <- r.svd$v[,1:nb.comp]
u <- r.svd$u[,1:nb.comp]
```

Using the first 10 right and left singular vector of the matrix we can reconstruct the image using:

```
img.compressed <- u %*% diag(d) %*% t(v)
img.compressed[img.compressed < 0] <- 0
img.compressed[img.compressed > 1] <- 1
```

```
plot.image(img.compressed,"10 components")
```



(a) Explore different number of component to see the effect on the compressed image. For example 10, 30, 50 and 640.

(b) Produce the image above with 100 components but using the colored photo.

(c) If you have time, explore the same procedure on a photo on your choice