Machine Learning - Coursework 1

- **Calculate and plot the average face of the training set, then write a function to find a PCA basis of size M, where the inputs will be M and X, the matrix containing the training set. Clearly describe all aspects of your function, then use it to plot the first 5 eigenfaces of the training set.**

To calculate the average face of the training set, I simply calculate the average pixel values for each pixel across the trainings set.

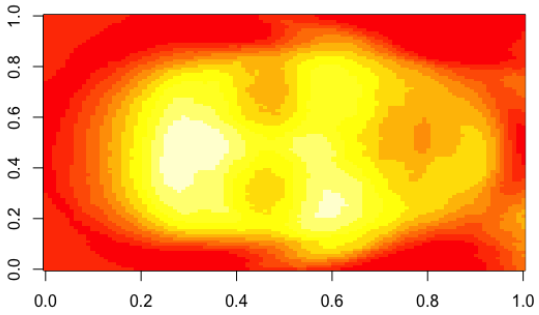
```
library(rARPACK)
library(philentropy)


# I load the raw csv's
faces.train.inputs <- read.csv("./2018_ML_Assessed_Coursework_1_Data/
                                Faces_Train_Inputs.csv",head=FALSE)
faces.train.label <- read.csv("./2018_ML_Assessed_Coursework_1_Data/
                               Faces_Train_Labels.csv",head=FALSE)
faces.test.inputs <- read.csv("./2018_ML_Assessed_Coursework_1_Data/
                               Faces_Test_Inputs.csv",head=FALSE)
faces.test.label <- read.csv("./2018_ML_Assessed_Coursework_1_Data/
                              Faces_Test_Labels.csv",head=FALSE)


# I turn the input values into a list of 320 matrices, each matrix a 112 x 92 value
# of pixels corresponding to each image .. I need to use lapply again on the
#result because apply gives the matrices in a weird form
faces.train.inputs.cleaned <- lapply(apply(X=faces.train.inputs,
                                           MARGIN=1,
                                           function(x) list(matrix(as.numeric(x),
                                           nrow = 112))), "[[", 1)

# Here I calculate the average face
avg.face <- Reduce('+', faces.train.inputs.cleaned) /
  length(faces.train.inputs.cleaned)
image(avg.face)
```

Figure 1: Average face of the training set



some more text

Figure 2: Average face of the training set