Classification with R

The theme of this tutorial is a different methods of classification with R. We will briefly go through following ways to classify a sample:

1. k-nearest neighbour
2. linear discrimination
3. quadratic discrimination
4. neural network

1 Train and test samples

You have to read 2 samples: the data with the training sample and the data you are going to classify

```
>df_train=read.csv("train.csv",header=TRUE)
>df_test=read.csv("test.csv",header=TRUE)
```

2 K-nearest neighbours

Load the package which contains knn method

```
>library(class)
>?knn
```

Try to classify your dataset:

```
k_res=knn(as.matrix(data.frame(df_train$x,df_train$y)),
          as.matrix(data.frame(df_test$x,df_test$y)),as.matrix(df_train$class),k=3)
```

In the case above you've used 3 nearst neighbours. Try to change `k` and compare result. Plot the train and test data.

2.1 Failure ratio

What is the ratio of missclassification? Try to do the test on the training dataset itself:

```
k_t=knn(as.matrix(data.frame(df_train$x,df_train$y)),
        as.matrix(data.frame(df_train$x,df_train$y)),as.matrix(df_train$class),k=3)
```

Compare results:

```
>comp=as.numeric(as.matrix(k_t))-as.numeric(df_train$class)
>i_wrong=length(abs(comp[comp!=0]))/length(comp)
>print(i_wrong)
```

Try to change number of neighbours and trace `i_wrong`. When the method start to fail? Plot different cases to get the answer.

3 Linear discrimination

Load library and create a classificator:

```
>library(MASS)
>lcl=lda(data.frame(df_train$x,df_train$y),df_train$class)
```

This classificator you can use on your data:

```
>cl=predict(lcl,df_test)
```

The result you will find in `cl$class`. Plot training dataset and test dataset with classification. Plot the discrimination line (see tasks to this werkcollege to find coefficients).

4 Quadratic discrimination

Repeat the section above, this time with `qda`

```
>lcl=qda(data.frame(df_train$x,df_train$y),df_train$class)
>cl=predict(lcl,df_test)
```

It is easy to draw a partitioning line with the use of package `klaR` (should be installed first!) Load package from the page of werkcollege and install it in your local Rlib directory:

```
virgo01>R CMD INSTALL -l /Users/user/<your_name>/R_libs klaR_0.6-3.tar.gz
```

Draw a plot:

```
> drawparti(cl$class,df_test$x,df_test$y,method="qda")
```

5 Neural network

Repeat the training and classification, this time with neural network of 5 elements in the hidden layer and linear output units

```
>library(nnet)
>ncl=nnet(data.frame(df_train$x,df_train$y),df_train$class,size=5,linout=TRUE)
>cl=round(predict(ncl,df_test)
```

Plot the result with the training set.