# Machine learning II
# Master Data Science
### Winter Semester 2023/24 — Prof. Tim Downie

# Workshop 7
# Support vector machines 2
## *with solutions*

There is a written exercise at the end of this worksheet.
The required packages for The R-exercises are `e1071`, `ISLR2`, `rpart`, `ROCR` and `MASS`.

**Exercise 1 Tutorial**
A good website explaining the subject of SVMs with out too much maths can be found here: *Support Vector Machines in R* Click here[1]

Included in this tutorial is a nicer plotting method than the default in the package `e1071`. If using more than 2 predictor variables then you need to take care defining the grid using the two variables you want to plot.

Read through this tutorial. **The second half includes R code which you should run yourself.** The data set can be downloaded directly from that web page; it originates from the more advanced course book for this course *Hastie, Tibshirani and Friedman*.

**Exercise 2 Non-linear SVMs: using different kernels**
Work through James Section 9.6.2 and 9.6.3 (pp. 392), which fits the the SVM algorithm using non-linear kernels. When you get to the `svm` command using the radial kernel function, fit SVM models with the following kernels:

- Linear kernel

  ```
  > svmfit<-svm(y~., data=dat[train,], kernel="linear", scale=TRUE, cost=1)
  > plot(svmfit, dat[train,])
  ```

  No boundary in the data region is found, and all the red points are misclassified. Varying the cost does not help.

[1] The link works in most PDF viewers but not all web browsers. If the link doesn't work, the URL is https://www.datacamp.com/community/tutorials/support-vector-machines-r

- Polynomial kernel with degree 2

  ```
  > svmfit<-svm(y~., data=dat[train,], kernel="polynomial",  degree=2,
  +  scale=TRUE, coef0=1, cost=0.1)
  > plot(svmfit, dat[train,])
  ```

  With a quadratic polynomial kernel two distinct borders are possible. Try increasing the cost using a few values between 1 and 10, and look at the effect on the boundary.

  Use `tune()` to find the best value for the cost parameter. and obtain the confusion matrix. This result is only very slightly worse than with the radial kernel in the book.

  **Remember** that the parameter cost is the reverse of the parameter $C$ in the lecture notes. A high value for `cost` penalises each support vector heavily. In the notes the parameter was similar to a budget and a high value allowed more support vectors.

- Try increasing the degree, and don't forget the arguments `scale=T` and `coef0=1`. The boundaries are slightly different but the results ar not any better.

**Note:** Last week you used the package `pROC` to obtain the ROC plots, which I prefer to `ROCR`. As James et al. uses `ROCR` this is the package used this week.

**Exercise 3  SVM with multiple classes**
Work through the short section 9.6.4. Which is a simple example of SVM with three classes. Although the support vector classifier algorithm is run 3 times, the code to do this is no more complex than for binary classification.

**Exercise 4  Using SVMs on a practical data set**
In Section 9.6.5 you will analyse the Khan *gene expression* data (in the ISLR package). This data set does not have many observations (63 in the training set) but 2308 variables each corresponding to a different gene. The value indicates how strongly each gene was "expressed". Many statistical learning methods have algorithmic problems when $p \gg n$, but this is not a problem with SVM.

# Exercise to do at home

Work through Exercise 3 in James et al. Section 9.7 page 398. This could be typical of an exam question on this subject.
N.B. In Part (g) the task is to find a hyperplane which is separating but not optimal, rather than any non-optimal hyperplane.

*(b) $\beta_0 = -0.5$, $\beta_1 = 1$ and $\beta_2 = -1$, i.e. $-0.5 + x_1 - x_2 = 0$ is one solution. Any multiple of these coefficients is sufficient.*

*(c) classify blue if $-0.5 + x_1 - x_2 \geqslant 0$ otherwise classify red. N.B. the direction of the inequality depends on the signs of $\beta_1$ and $\beta_2$.*

*(d) The value of $M$ is $\frac{\sqrt{2}}{4}$ which is half the width of the margin. (e) Observations 2,3,5 and 6 are the support vectors.*

*(f) As the seventh point is not a support vector, a small change will not move it into the margin and so will not change the maximal margin.*

*(g) an easy solution is $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = -1$, i.e. $x_1 - x_2 = 0$.*

*(h) Any point on the wrong side will do, e.g. (1,3) Blue.*