Workshop 5
Introduction to support vector machines

In this workshop you will need the packages `ISLR`, `MASS`, `gbm` and `e1071`. Load these packages, installing them if need be.

Exercise 1 James Lab

The package to fit SVM models in R is called `e1071`, an informative name which originates from an internal department code at Vienna University!

Linear Support vector classifier

Work through subsection **9.6.1** in James et al. on page 359. Read the text carefully! Notice in particular, that the cost is defined differently as in the notes, so a large cost gives a small margin.

Exercise 2 Birth weight data set

In the `MASS` library, there is a data set called `birthwt`. Read the help page for this data set. The variable `low` is a binary variable for low birth weight. You will use SVM to investigate a classifier for `low`.

Why is it not sensible to use the variable `bwt` in this classifier?

`svm` expects a factor variable for the outcome variables so convert `low`

```
> birthwt$low<-as.factor(birthwt$low)
```

Fit an SVM for `low` dependent on all variables except for `bwt`. Start with a cost 0.1 and don't forget the option `kernel="linear"`. Obtain a classification matrix for this model, and calculate the misclassification rate.

Increase the cost and investigate the effect on the misclassification rate.

Does scaling the data have a noticeable effect on the model results?