Support vector machines

Tasks (Lab 5):

1. Generate 2 synthetic datasets (**scored task**):

   (a) Dataset 1:

   - Consider two classes $y = 1$ and $y = 0$. For each class, there are 500 observations.
   - Consider $p = 2$ features: x_1 and x_2.
   - Let x_1 and x_2 in class $y = 0$ be generated independently from $N(0, 1)$.
   - Let x_1 and x_2 in class $y = 1$ be generated independently from $N(m, 1)$, where m is a parameter. Consider e.g. $m = 0.5, 1, 2, 3$. The larger the m, the easier is the classification problem.

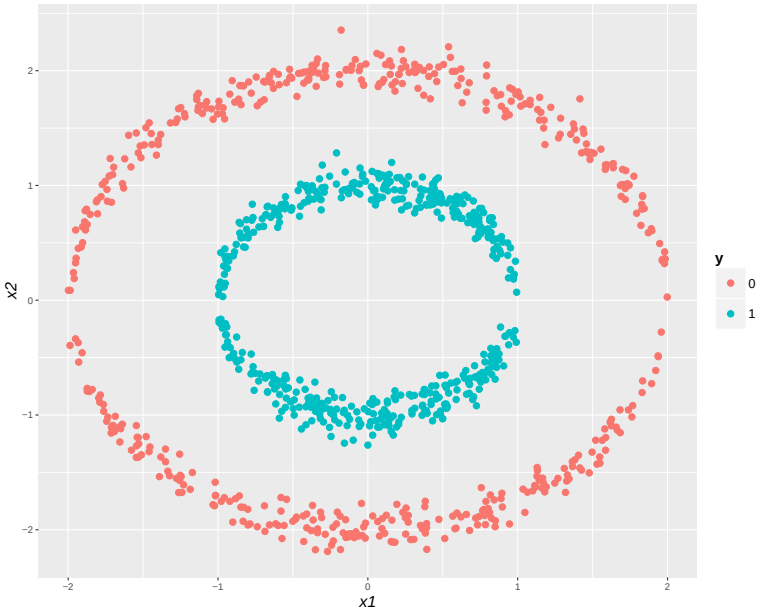
   Example data (for $m = 2$):

(b) Dataset 2:

Generate artificial data as follows:

- Consider two classes $y = 1$ and $y = 0$. For each class, there are 500 observations.
- Consider $p = 2$ features: x_1 and x_2.
- Let $x_1^2 + x_2^2 = 1$, in class $y = 1$, where $x_1 \sim U[-1, 1]$.
- Let $x_1^2 + x_2^2 = 4$, in class $y = 0$, where $x_1 \sim U[-2, 2]$.
- Add some noise to make a classification problem more difficult.

Example data:



2. Perform the following steps (**scored task**)

- Run SVM method on the Dataset 1 and Dataset 2 using different kernel functions: linear, radial, polynomial.
- Visualize the results. Use e.g. different colours or symbols for predictions.
- Investigate the influence of parameter m (in the case of Dataset 1) and variance of the noise (in the case of Dataset 2). Try 4 different values of the variance of the noise.