Circle Dataset

1 Introduction

The circle dataset is a synthetic dataset from `mlbench`. The circle problem's inputs are uniformly distributed across the `d`-dimensional cube with corners. This is basically a 2-class classification problem: the first class is a `d`-dimensional ball in the centre of the cube, while the remainder forms the second class. The ball size is selected so that both classes have an identical prior probability of 0.5.

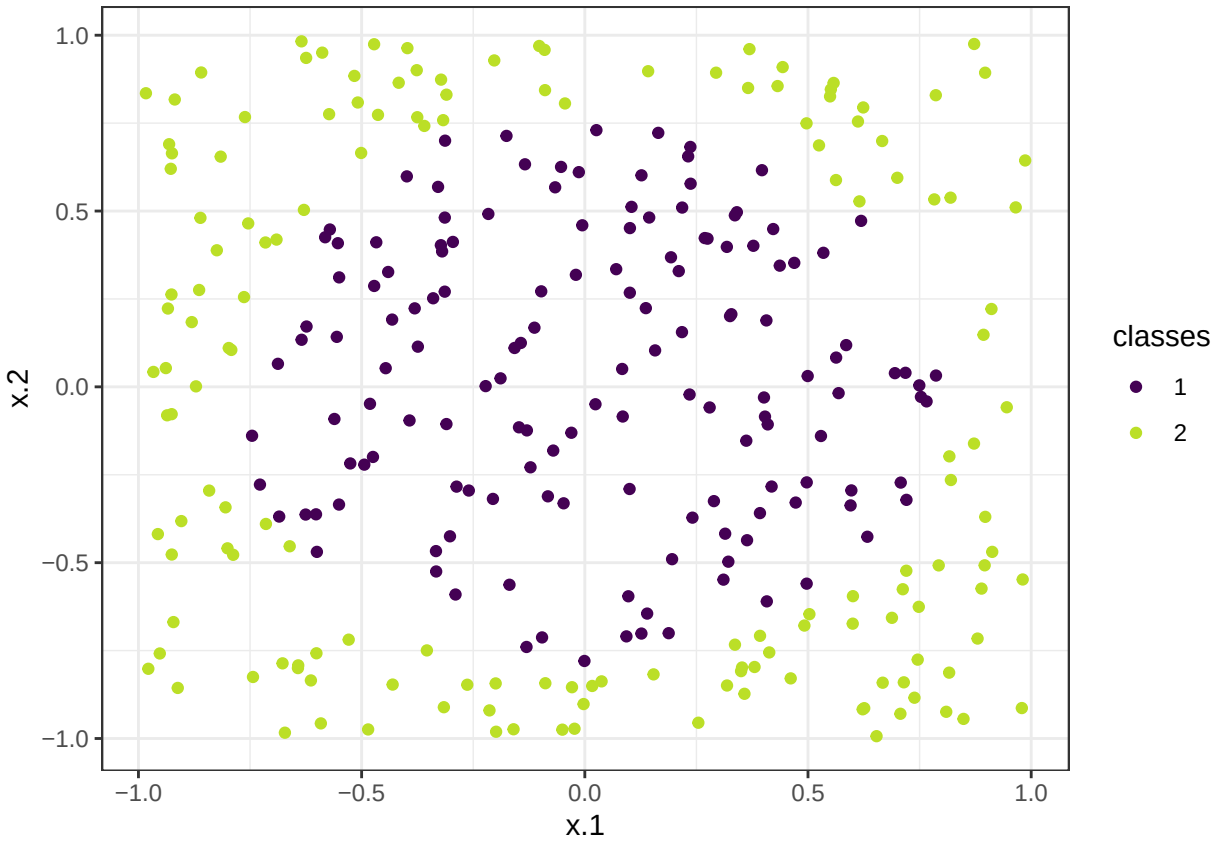
To generate the dataset, we need to define two parameters: `n` - the sample size and `d` - the number of dimensions. Below is an example of the problem with 2 dimensions.

```
# load the dataset from mlbench
circle_2d <- mlbench.circle(n = 300, d = 2) %>% as_tibble()
print(circle_2d)
```

```
## # A tibble: 300 x 3
##        x.1     x.2 classes
##      <dbl>   <dbl> <fct>
##  1 -0.926  -0.0776 2
##  2  0.848  -0.944  2
##  3 -0.199  -0.981  2
##  4  0.380  -0.797  2
##  5 -0.0444  0.806  2
##  6 -0.310   0.831  2
##  7  0.157   0.104  1
##  8  0.461  -0.829  2
##  9  0.362  -0.153  1
## 10 -0.591  -0.957  2
## # ... with 290 more rows
```

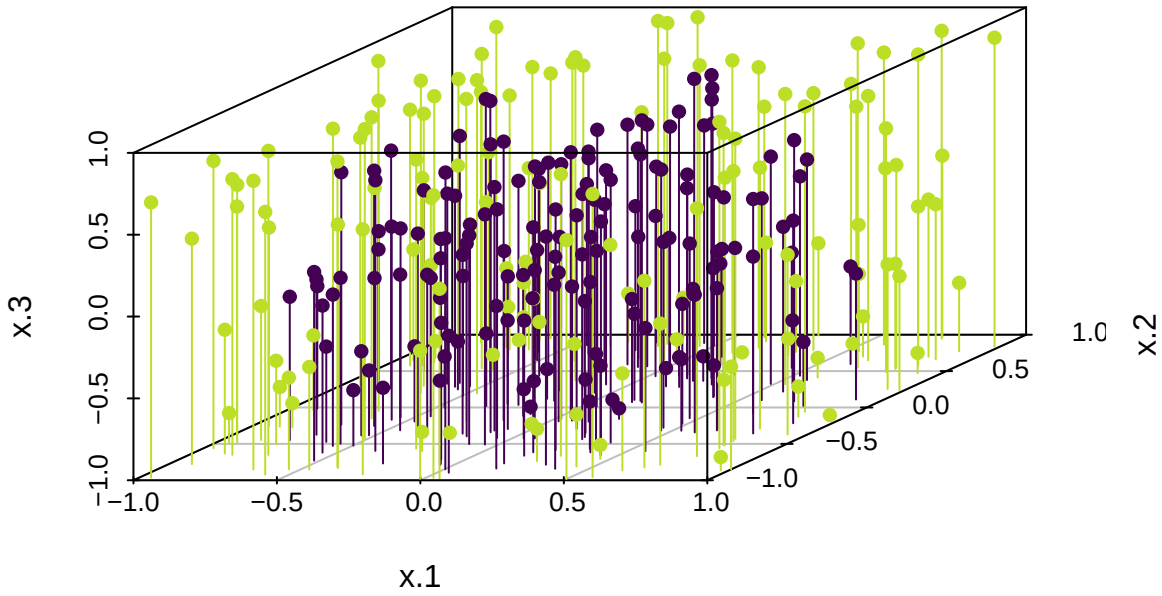
As we define it to be a 2-d problem, it has 2 features/coordinates and a target named `classes`. Let's visualize this 2-d dataset with `ggplot2`:

```
viridis_2 <- viridis(2, end = 0.9)
ggplot(data = circle_2d, mapping = aes(color = classes, x = x.1, y = x.2)) +
        geom_point() +
        scale_color_manual(values = viridis_2)
```



We can also easily create and visualize the circle dataset at 3 dimensions:

```
circle_3d <- mlbench.circle(n = 300, d = 3) %>% as_tibble()
scatterplot3d(
        circle_3d[,1:3],
        pch = 16,
        type = "h",
        color = viridis_2[as.numeric(circle_3d$classes)]
)
```


x.3
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
x.1
1.0
0.5
0.0
−0.5
−1.0
x.2