

Spirals Dataset

1 Introduction

The spirals dataset is a synthetic dataset from `mlbench`. The spirals problem's inputs are points on two entangled spirals. If `sd` is greater than zero, Gaussian noise is introduced to each data point.

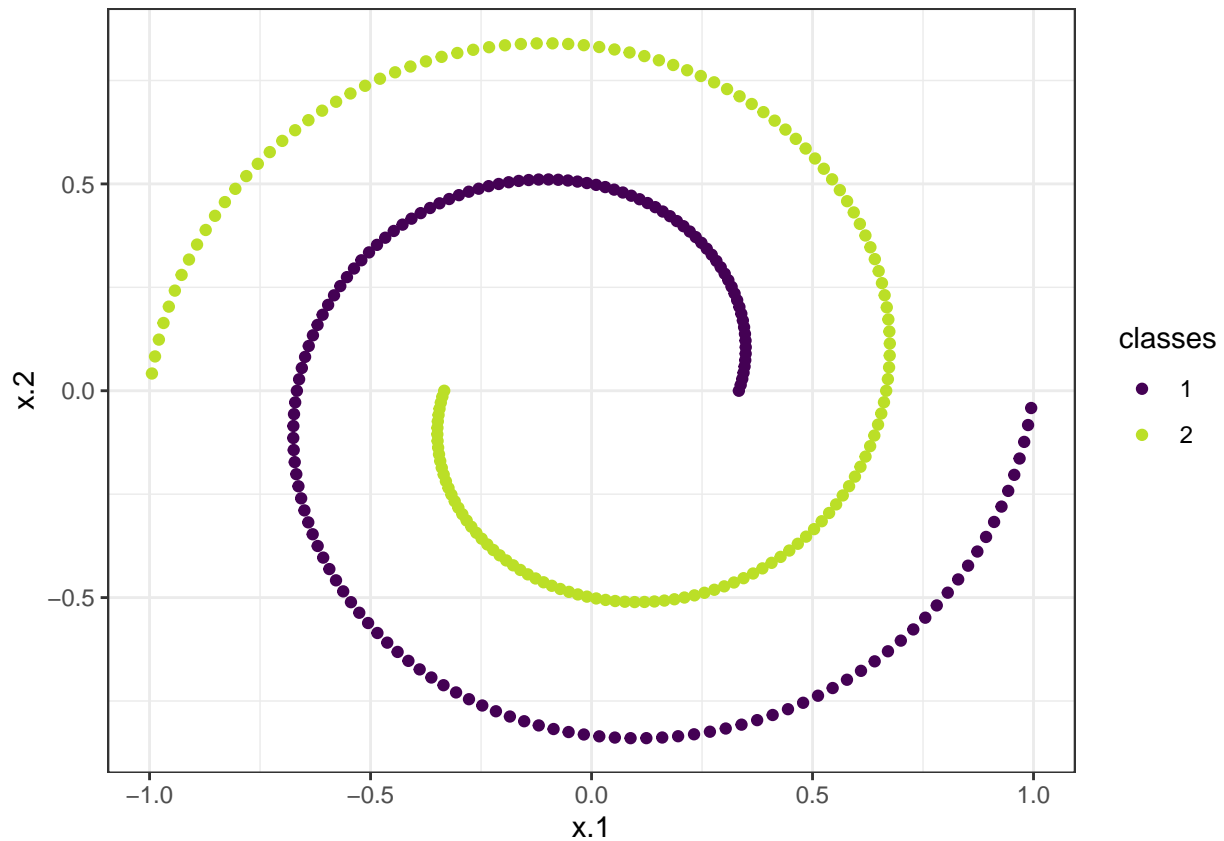
To generate the dataset, we need to define `n` - number of patterns to create, `cycles` - the number of cycles each spiral makes, and `sd` - standard deviation of data points around the spirals. Below is an example of the spirals dataset with one cycle each (`cycles = 1`) and no noise (`sd = 0`):

```
# load the dataset from mlbench
spirals_1 <- mlbench.spirals(n = 300, cycles = 1, sd = 0) %>% as_tibble()
print(spirals_1)
```

```
## # A tibble: 300 x 3
##       x.1      x.2 classes
##   <dbl>   <dbl> <fct>
## 1 -0.292 -0.298  2
## 2  0.333  0      1
## 3  0.0318 -0.506  2
## 4  0.277  0.746  2
## 5  0.337  0.0141 1
## 6  0.672  0.172  2
## 7  0.341  0.0286 1
## 8  0.344  0.0434 1
## 9  0.675  0.114  2
## 10 0.0534 -0.508  2
## # ... with 290 more rows
```

Basically, the spirals dataset will always have 2 features `x.1` and `x.2` indicating the 2-d coordinates of the data points and one target `classes` indicating which class each data point belongs to. Because there is no added noise, the data points from each spiral will align perfectly on the curve.

```
viridis_2 <- viridis(2, end = 0.9)
ggplot(data = spirals_1, mapping = aes(color = classes, x = x.1, y = x.2)) +
  geom_point() +
  scale_color_manual(values = viridis_2)
```



We can also easily generate another spirals dataset with a more complex parameters: 2 cycles and `sd = 0.1`:

```
spirals_2 <- mlbench.spirals(n = 300, cycles = 2, sd = 0.1) %>% as_tibble()
ggplot(data = spirals_2, mapping = aes(color = classes, x = x.1, y = x.2)) +
  geom_point() +
  scale_color_manual(values = viridis_2)
```

