Exercise1

**data\_blobs**

x1\_test x1\_train x2\_test x2\_train y\_test y\_train

0 0.982570 -1.643388 0.982570 7.526920 0.0 3.0

1 -1.378426 -1.149899 -1.378426 7.893253 3.0 3.0

2 -1.379980 -2.632746 -1.379980 2.631098 3.0 2.0

3 -1.757832 -1.107830 -1.757832 2.920145 2.0 2.0

4 -1.186530 -2.783666 -1.186530 2.992115 2.0 2.0

.. ... ... ... ... ... ...

235 NaN -1.558767 NaN 7.248162 NaN 3.0

236 NaN 1.379647 NaN 4.548264 NaN 0.0

237 NaN -1.131214 NaN 6.766522 NaN 3.0

238 NaN -1.958667 NaN 2.430086 NaN 2.0

239 NaN 0.529417 NaN 3.807924 NaN 0.0

**Accuracy for blobs dataset with k=15: 0.05**

**data\_moons**

x1\_test x1\_train x2\_test x2\_train y\_test y\_train

0 0.114329 1.505058 0.114329 0.421019 1.0 1.0

1 -0.431282 0.974202 -0.431282 -0.228465 0.0 1.0

2 0.783993 -0.001829 0.783993 -0.334952 0.0 1.0

3 1.918391 0.973972 1.918391 0.376482 1.0 0.0

4 0.551403 0.837592 0.551403 0.098647 1.0 0.0

.. ... ... ... ... ... ...

235 NaN 1.223116 NaN -0.697216 NaN 1.0

236 NaN 1.052601 NaN 0.908250 NaN 0.0

237 NaN -1.422420 NaN 0.185039 NaN 0.0

238 NaN -0.157485 NaN 0.778762 NaN 1.0

239 NaN 0.461403 NaN -0.101698 NaN 0.0

**Accuracy for moons dataset with k=15: 0.14**

**data\_full**

x1\_test x1\_train x2\_test x2\_train y\_test y\_train

0 0.973055 0.789789 0.973055 0.498613 1.0 1.0

1 0.274442 0.326297 0.274442 0.280568 0.0 1.0

2 0.278993 0.265568 0.278993 0.882993 0.0 0.0

3 0.558535 0.331562 0.558535 0.990420 1.0 0.0

4 0.462017 0.559319 0.462017 0.487208 0.0 1.0

... ... ... ... ... ... ...

2395 NaN 0.072187 NaN 0.458168 NaN 0.0

2396 NaN 0.374616 NaN 0.557913 NaN 0.0

2397 NaN 0.127180 NaN 0.707989 NaN 0.0

2398 NaN 0.992417 NaN 0.295864 NaN 1.0

2399 NaN 0.751318 NaN 0.567816 NaN 1.0

**Accuracy for full dataset with k=15: 0.13**

图表, 散点图

描述已自动生成

Exercise 2

2(e)

If the bandwidth parameter h is small, it's like zooming in really close. We can see every little detail, but sometimes that makes the picture too busy because you're also seeing all the little specks of dust and scratches. In terms of the data, this means the curve might Z-shaped a lot, trying to hit every point, even the noise.

If h is big, it's like zooming out for a wide shot. Now, we can't see all the tiny details, and everything looks smooth. For your data, this means the curve is smooth and gentle, without all the Z-shaped, but it might miss out on important patterns.