A brief introduction to the k-nearest neighbors classifier

Course: English for Academic Purposes

Student: Eduardo Henrique Basilio de Carvalho

Universidade Federal de Minas Gerais, May 13, 2025

Table of Contents I

Table of Contents II

Problem introduction

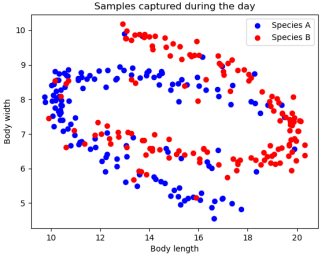
- Two species
- Count sightings of each
- Take some measurements

Training data

- Species are distinguishable by fur color
- Measure body length and width with a camera

Day measurements

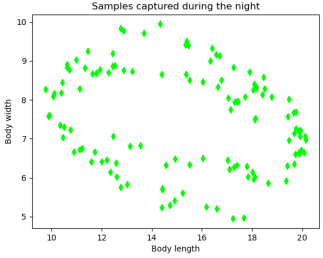
Figure: Day sightings plot



UFMG

Night measurements

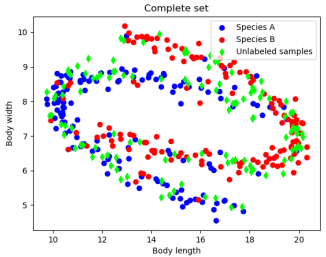
Figure: Night sightings plot



UFMG

Data superposition

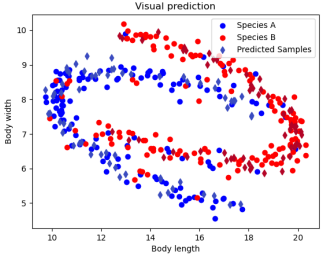
Figure: Superimposed sightings plot



UFMG

Visual prediction

Figure: Visually predicted samples



UFMG

Higher dimensional data I

Table: Four-dimensional train samples

Sample	Feature 0	Feature 1	Feature 2	Feature 3	Label
0	-1.12	0.43	-1.5	0.55	1
1	1.93	-1.71	-0.75	-1.15	0
2	1.7	1.63	1.44	-0.42	1
3	-2.45	0.64	-0.48	0.17	1
4	1.14	-0.56	0.46	-1.04	1
5	-1.29	-1.58	-0.04	-2.11	0
6	-1.56	-1.13	-1.08	0.7	0
7	2.02	-0.14	-1.25	-1.96	1
8	1.37	0.01	-3.05	1.66	0

Higher dimensional data II

Table: Four-dimensional test sample

Feature 0	Feature 1	Feature 2	Feature 3	Label
-0.72	-0.41	1.21	-2.49	?

Distance

Table: Four-dimensional train samples with distances

Sample	Feature 0	Feature 1	Feature 2	Feature 3	Label	Distance
0	-1.12	0.43	-1.5	0.55	1	1.62
1	1.93	-1.71	-0.75	-1.15	0	4.47
2	1.7	1.63	1.44	-0.42	1	5.24
3	-2.45	0.64	-0.48	0.17	1	4.73
4	1.14	-0.56	0.46	-1.04	1	6.04
5	-1.29	-1.58	-0.04	-2.11	0	6.87
6	-1.56	-1.13	-1.08	0.7	0	7.34
7	2.02	-0.14	-1.25	-1.96	1	8.29
8	1.37	0.01	-3.05	1.66	0	8.99

Closeness ranking

Table: Four-dimensional train samples ranked by distances

Sample	Feature 0	Feature 1	Feature 2	Feature 3	Label	Distance	Rank
0	-1.12	0.43	-1.5	0.55	1	1.62	1
1	1.93	-1.71	-0.75	-1.15	0	4.47	2
2	1.7	1.63	1.44	-0.42	1	5.24	4
3	-2.45	0.64	-0.48	0.17	1	4.73	3
4	1.14	-0.56	0.46	-1.04	1	6.04	5
5	-1.29	-1.58	-0.04	-2.11	0	6.87	6
6	-1.56	-1.13	-1.08	0.7	0	7.34	7
7	2.02	-0.14	-1.25	-1.96	1	8.29	8
8	1.37	0.01	-3.05	1.66	0	8.99	9

Nearest neighbor label

Table: Four-dimensional test sample labelled by its nearest neighbor

Feature 0	Feature 1	Feature 2	Feature 3	Label
-0.72	-0.41	1.21	-2.49	1

2D set recap

Edges to the nearest neighbor

Nearest neighbor prediction

Decision boundary

Prediction for 10-NN

Decision boundary for 10-NN

Dataset summary

Results

Questions?

Thank you! Questions?