

UNIVERSITÉ DE GENÈVE

DATA MINING

TP 1: NB

Author: Joao Filipe Costa da Quinta

E-mail: Joao.Costa@etu.unige.ch

March 29, 2021



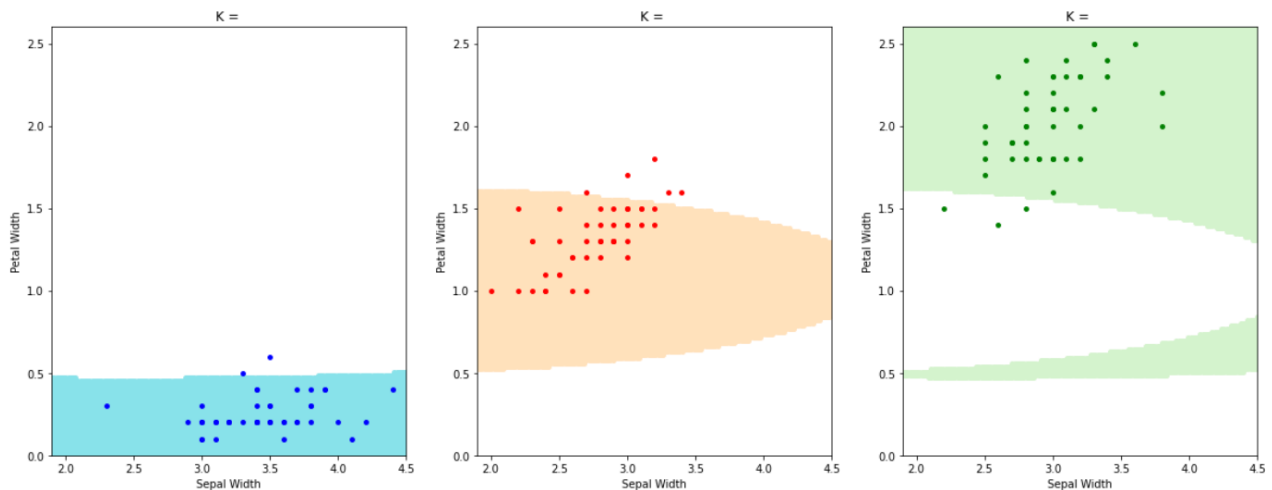
**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES
Département d'informatique

Continuous Case

For the continuous problem, I had an accuracy of 94% it is quite high, with a bit more of training data, we would probably get a better accuracy. for our problem I feel like it is acceptable, however, if human life was at stake, maybe 94% wouldn't be high enough.

I set the same axes_lim for all three graphs:



And I found that the graph was much easier to analyse this way, we can clearly see the training data is for the most part separated, just a bit of overlap, this overlap does bring down our accuracy rate, since some data points that are represented in the green color, were predicted as the orange color.

We can also visualise each line that separates the classes.

What was unexpected, is the second part of data that was predicted as green, this should have been most likely blue, this probably brings our accuracy down.

Aside from this small mistake, all the fields are continuous.

Discrete Case

For the discrete case, I had around 80%, this is probably due to the fact that we suppose that the attributes are independent, when in fact they aren't.

Adding another dimension of information would help too, the location at time of impact, would be a good argument to add, as it is harder to survive if you are in a bellow level, rather than next to the life boats.