Question Exam IRIS

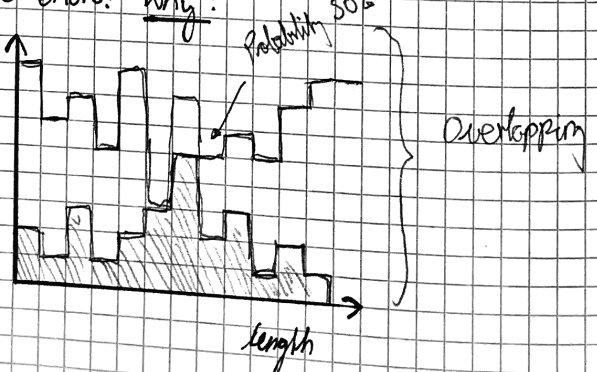
Given that an IRIS Sentosa has a sepal length in the range of around 4,2 - 6 cm and a sepal with of around 2,8 - 4 cm discuss the limitations of histogram as a method to classify the flowers.

- Only one feature!
- Histogram → Frequency + Buckets! → binning the intervals

Building a histogram involves counting which is a discrete process and is not a natural fit to the length of flowers we see in nature. Say we ~~divide~~ divide the sepal length range between 4,2 to 6 cm into bins we encounter the problem that we don't know yet how many bins to split the range into and yet we don't know the boundaries either and this may require a lot of experimentation with the actual data. Even if we do it, the sepal with is within the range of values for the sepal length and as such it seems very unlikely that the histogram alone can be used to unambiguously discriminate between the three flower categories.

↳ How could I build as fast a good Classification?

- We know that the binning and one feature alone will have some errors! Why?



Verdict: No single treshold value of the length will serve as a good discriminate between the two categories.

Solution:

Hypothesis Assumtions, means work with assumtions such as $P(\omega_1) = P(\omega_2)$ and ~~test~~ test them.

- Build a distribution of your data to form a decision.
- Define what you want to count!