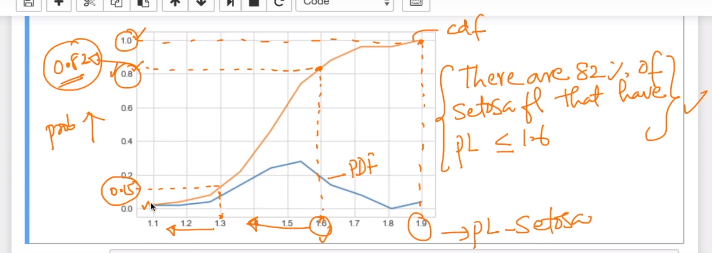
CDF (Cumulative Distribution Function) gives us the percentage of population present upto particular point,

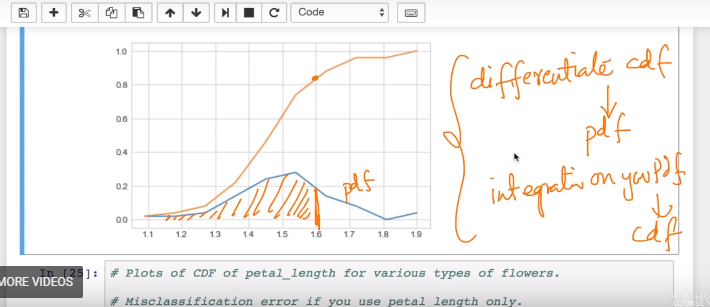
Ex: if we want to find the % age of set of flowers that are less than 1.6 length, then we find CDF of it, as for below graph it is 0.82 ie 82%

We can’t find this with PDF because it tells only how much % of population within a range.



**Note: Differentiation of CDF is PDF**

**Integration of PDF is CDF**



**Why do we actually use CDF in real world:**

If you are building a simple model using if else, with one variable then it will hep you know the % of accuracy and % of wrong.

For example let’s say our basic model on the basis of PDF is:

(remember versicular and virginica PDF’s are intersecting at 5)

If PL > 2 and PL < 5

Then it’s versicolor.

If PL > 2 and PL > 5

Then it’s virginica.

Now for these model in below graph we can see that for versicolor the CDF at length 5 (here 5 is intersecting length of versicolor and virginica PDF) is 0.95 that means 95% flowers are there upto length 5, so what does it concludes that our model is 95% accurate and making 5% mistake for remaining versicolor flower that are after 5 length.

Again for virginica the CDF at length 5 is 0.10 that is only 10% of virginica flowers are there upto length 5, so what does it concludes that our model is making mistake of 10% for classifying virginica and have accuracy of 90% for classifying virginica.

**This is the biggest usage of CDF for identifying accuracy if we are doing analysis on single variable.**

