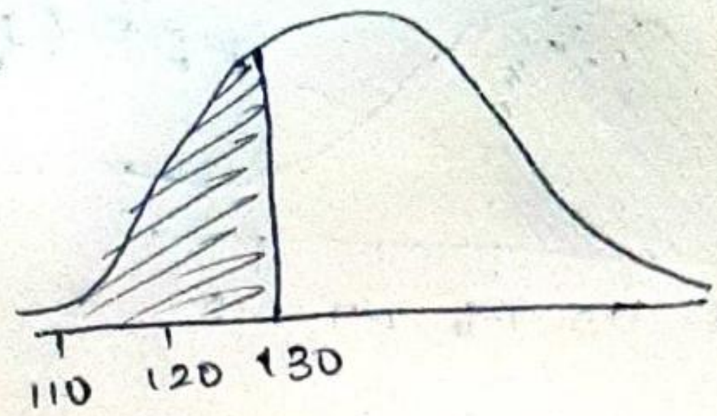
cumulative Density function (CDF):-

* cumulative density function can be defined as $F(x) = P(X \leq x)$.

* Also defined from $-\infty$ to x, it gives the values less then or equal to the x.



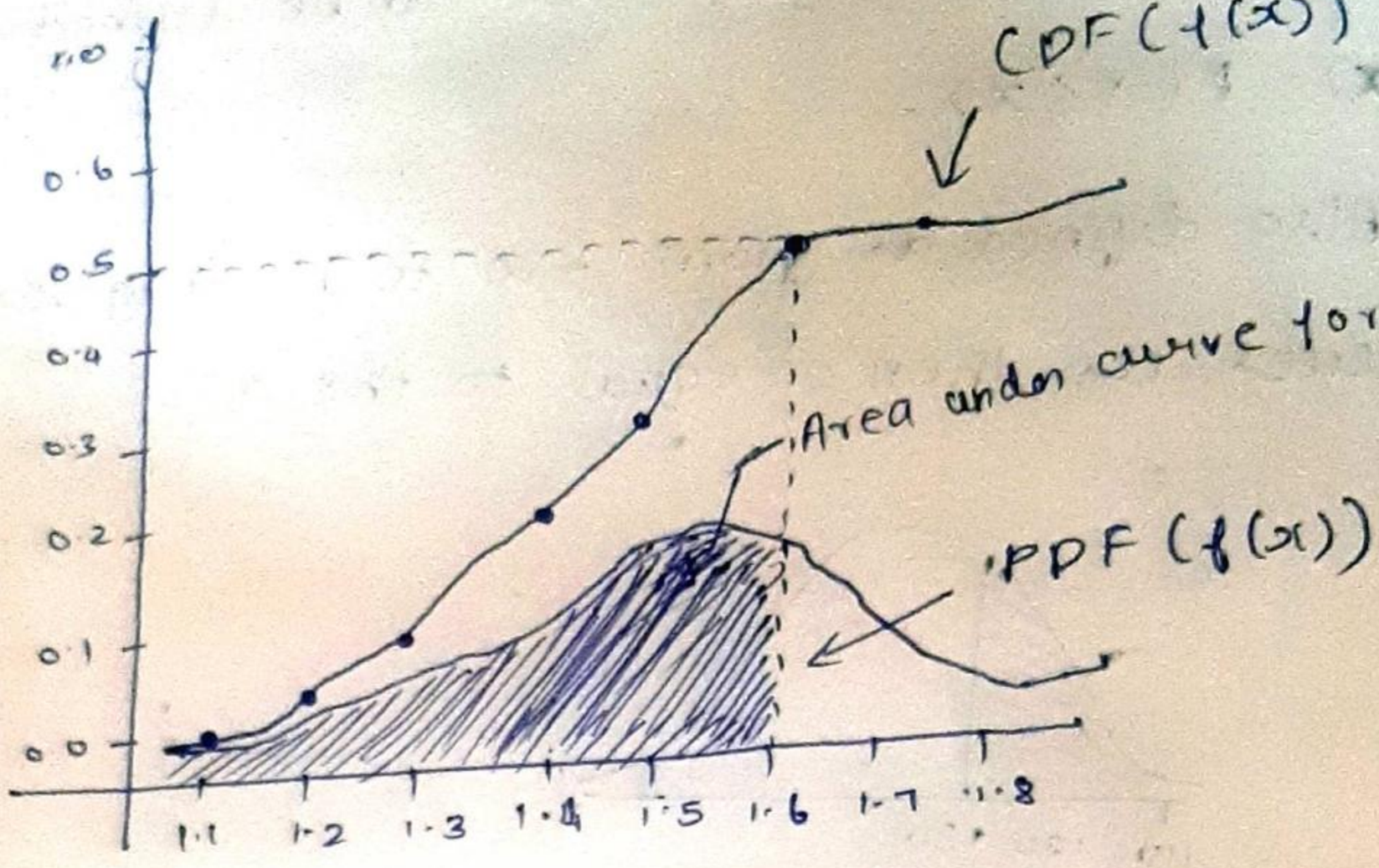
* If I ask find probability of $P(X \leq 130)$ it will basically add 110 and 120 till the 130, so the basic definition of cumulative, that is adding the consecutive values is defined.

Why CDF is used what it is doing in EDA?

* usually PDF shows how the data is distributed and with PDF we can find probability distribution within the intervals.

* The drawback with PDF is we cant directly analyse the distribution of the data by seeing the graph, everytime we need to calculate the area.

* This drawback is resolved by CDF in EDA

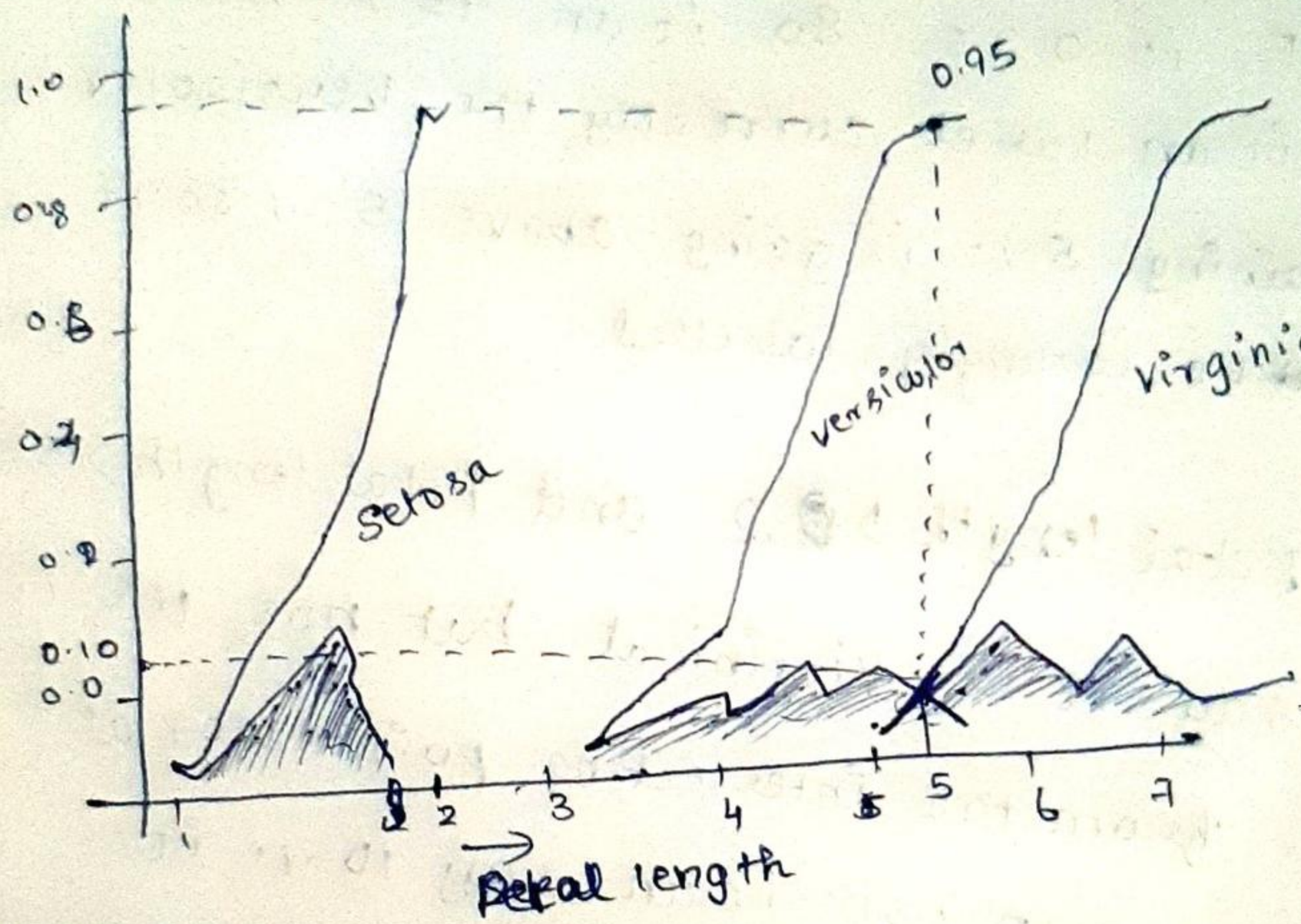


* The CDF will be derived from PDF
* But the doubt is how the CDF is built in association with PDF.
* The plotting points in CDF is generated from the value of area under the curve
* For ex. if $P(X \leq 1.1)$ is 0.01, that is marked straight to 1.1, if $P(X \leq 1.2)$ is 0.05, that is marked straight to 1.2 like wis if $P(X \leq 1.6)$ is 0.5, it is marked $P(X \leq 1.6)$ will add up all the values such as $P(X \leq 1.1) + P(X \leq 1.2) + P(X \leq 1.3) + P(X \leq 1.4) + P(X \leq 1.5) + P(X \leq 1.6)$
* This is how the CDF is built from PDF.
* Now we can easily analyse the distribution proportion just by seeing instead of calculating

* If u differentiate CDF we will get PDF, If u integrate PDF we will get CDF.

* CDF is the non-decreasing function, if the distribution start to fall, the CDF will have horizontal line.

Important Application why CDF is used :-



* It is the distribution of Petal length in Iris dataset.

* we can clearly see that the PDFs of Versicolor and Virginca was intersecting, now how can i fix the distribution range of Petal length.

* Another problem is how i can label the flowers accurately if the PDFs are intersecting.

Here CDF comes to action

* If Petal length < 2 ~~i~~ i can 100.1. correctly label setosa

If Petal length > 2 and Petal length ≤ 5 i can say it is Versicolor but not 100.1. because, from the intersecting point 100% the CDF, it 0.95. So i can 95.1. sure that i can label correctly the, Versicolor. Remaining 5.1. is going above 5, so it may ~~be~~ be wrongly labelled.

If Petal length > 2 and Petal length > 5 i can say it is Virginical but not 100%. because, from the intersecting point 100% at the CDF is 0.10 i can say 10.1. it may go wrong because it lies below 5, 90.1. i can label correctly.

This how CDF is used.