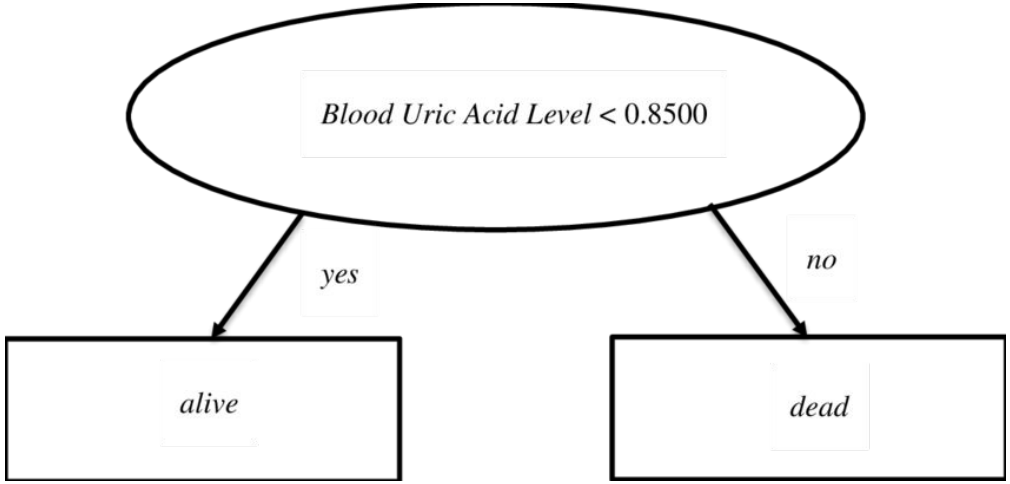
What is Decision tree Stump?

A decision stump is a one-level decision tree that makes a decision using a single feature and a single threshold.



Bagging and Boosting

Bagging (Bootstrap Aggregating)

Bagging trains many models independently on different random samples of the data, then averages or votes their predictions to reduce variance and overfitting.

Boosting

Boosting trains models sequentially, where each new model focuses more on the previous model's mistakes, and then combines them to reduce bias and improve accuracy.

What is Weak learner?

Weak learner

A weak learner is a model that performs only slightly better than random guessing, but is simple and fast.

Key idea:

It is not very accurate, but it learns some useful pattern.

Strong Learner

Strong learner

A strong learner is a model that achieves high accuracy and captures complex patterns.

Key idea:

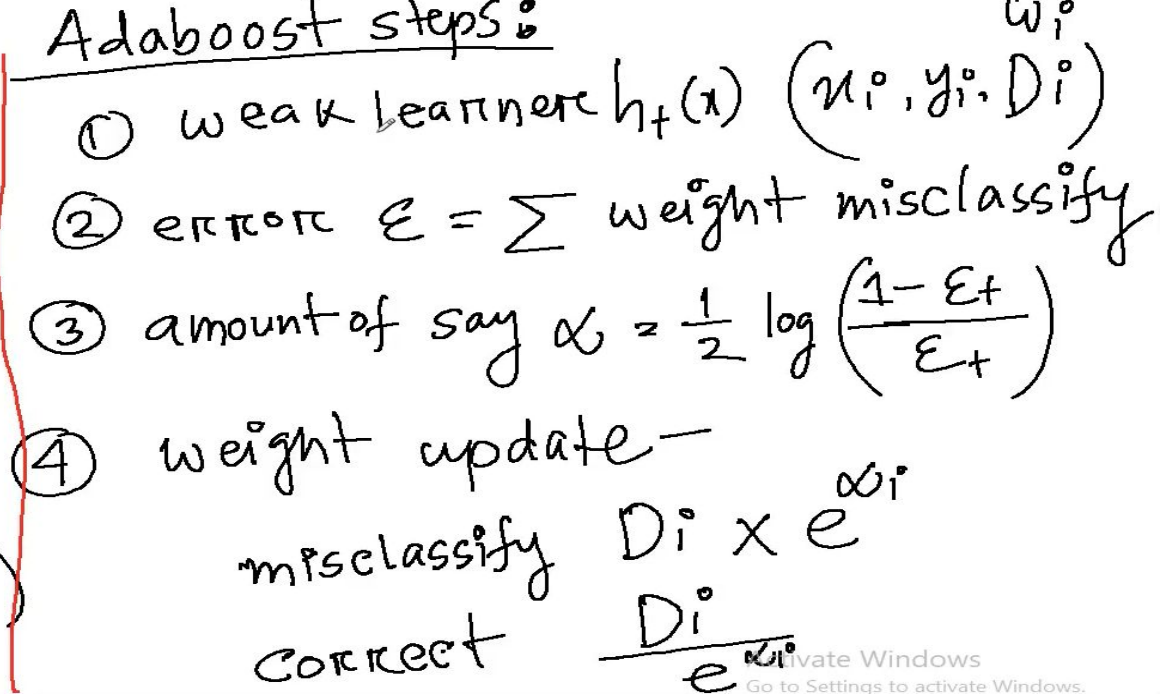
It performs well on its own.

Differences between Bagging and Boosting

1. Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types.
2. **Bagging aims to decrease variance**, not bias while Boosting aims to decrease bias, not variance.
3. In Baggiing each model receives equal weight whereas in Boosting models are weighted according to their performance.
4. In Bagging each model is built independently whereas in Boosting new models are influenced by performance of previously built models.
5. In Bagging different training data subsets are randomly drawn with replacement from the entire training dataset. In Boosting every new subsets contains the elements that were misclassified by previous models.
6. Bagging tries to solve over-fitting problem while Boosting tries to reduce bias.
7. If the classifier is unstable (high variance), then we should apply Bagging. If the classifier is stable and simple (high bias) then we should apply Boosting.

Now Understand AdaBoost

Steps



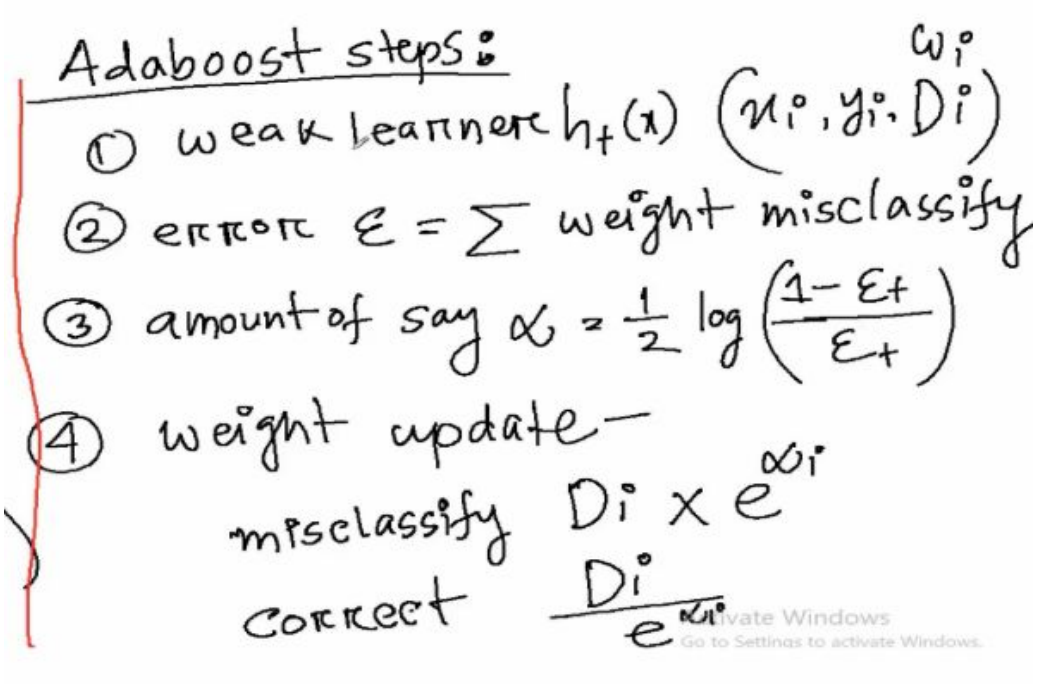
amount of say is the weight assigned t
weak learner, reflecting how much
influence it has in the final prediction.

Chest pain	Block arteries	weight	Heart disease
Yes	Yes	205	yes
NO	Yes	180	Yes
Yes	NO	210	Yes
Yes	Yes	167	Yes
NO	Yes	156	No
NO	Yes	125	NO
Yes	NO	168	NO

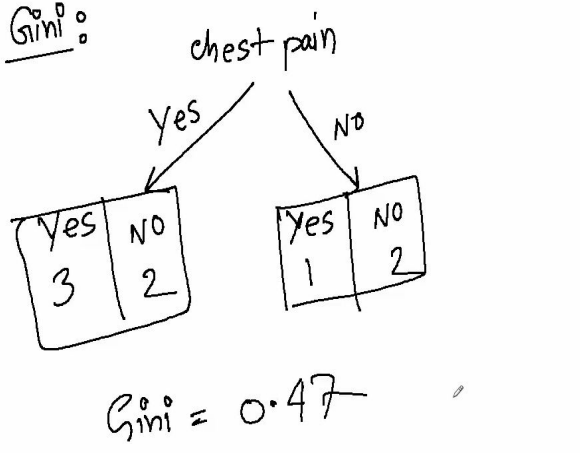
Add weight , rules 1/(total number of samples)

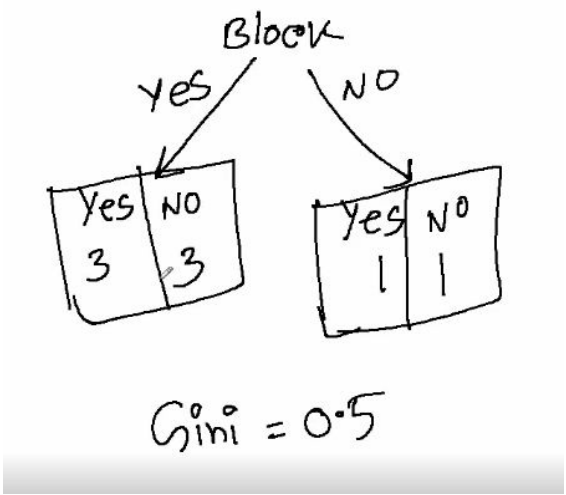
Chest pain	Block arteries	weight	Heart disease	weight
Yes	Yes	205	Yes	$\frac{1}{8}$
NO	Yes	180	Yes	$\frac{1}{8}$
Yes	NO	210	Yes	$\frac{1}{8}$
Yes	Yes	167	Yes	$\frac{1}{8}$
NO	Yes	156	No	$\frac{1}{8}$
NO	Yes	125	NO	$\frac{1}{8}$
Yes	NO	168	NO	$\frac{1}{8}$
Yes	Yes	172	NO	$\frac{1}{8}$

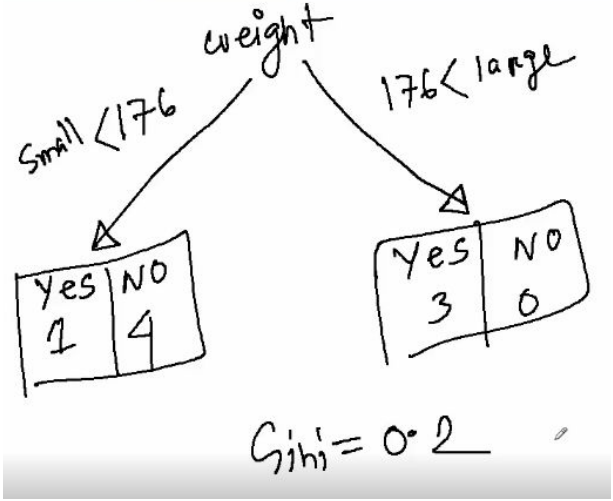
Steps


Adaboost steps:
① weak learner $h_t(x)$ (x_i, y_i, D_i) w_i
② error $\varepsilon = \sum$ weight misclassify
③ amount of say $\alpha = \frac{1}{2} \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$
④ weight update –
misclassify $D_i \times e^{\alpha_i}$
correct $\frac{D_i}{e^{\alpha_i}}$
Activate Windows
Go to Settings to activate Windows.

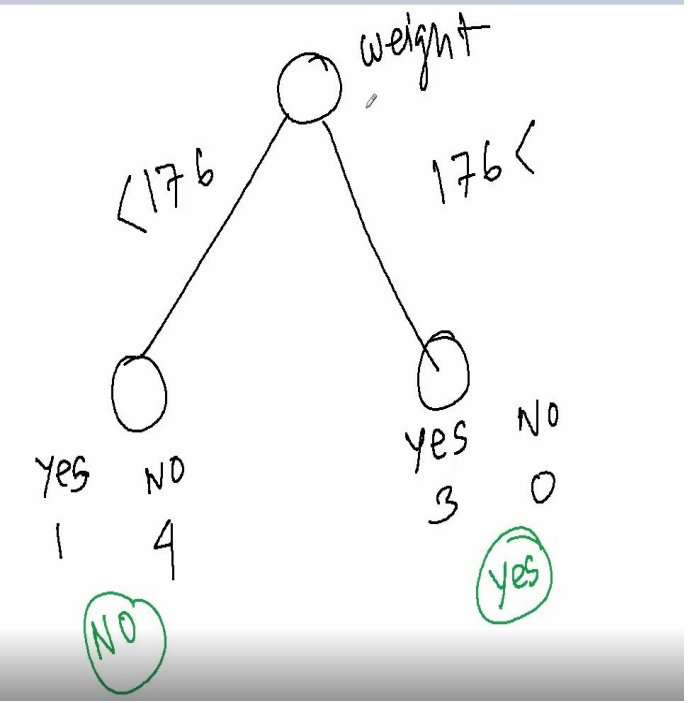
Split Decision tree based on Gini value

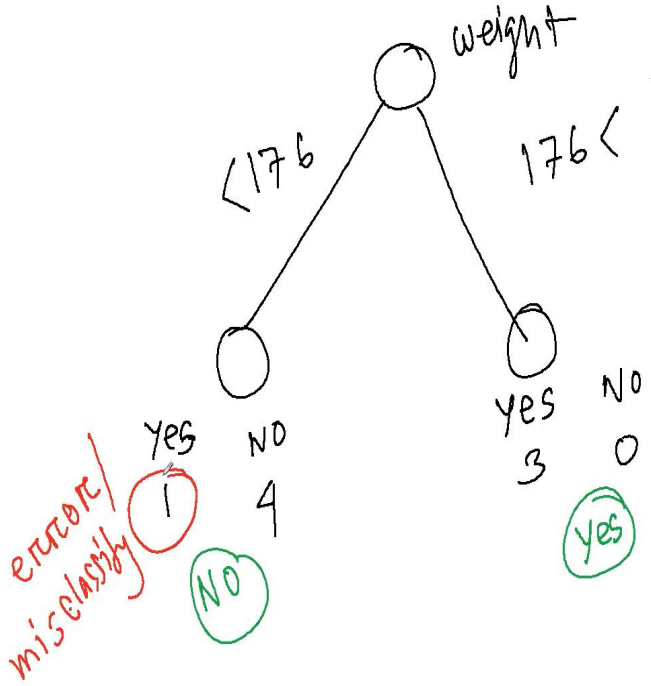






Step 2 : Error Calculation





$$\varepsilon = \sum 1 \times \frac{1}{8} = \frac{1}{8}$$

Step 3 : Error Calculation

$$\alpha = \frac{1}{2} \log \left(\frac{1 - \frac{1}{8}}{\frac{1}{8}} \right)$$

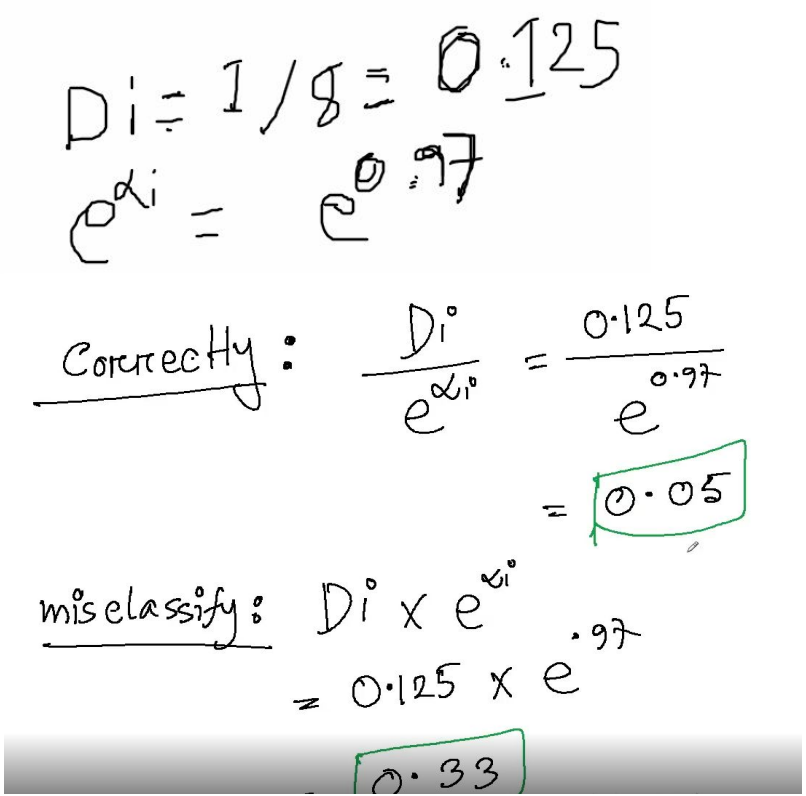
$$= \frac{1}{2} \log \left(\frac{\frac{7}{8}}{\frac{1}{8}} \right)$$

$$= \frac{1}{2} \log \left(\frac{7}{8} \times \frac{8}{1} \right)$$

$$= \frac{1}{2} \log_e (7)$$

$$= 0.97$$

Step 4 : Weight Update



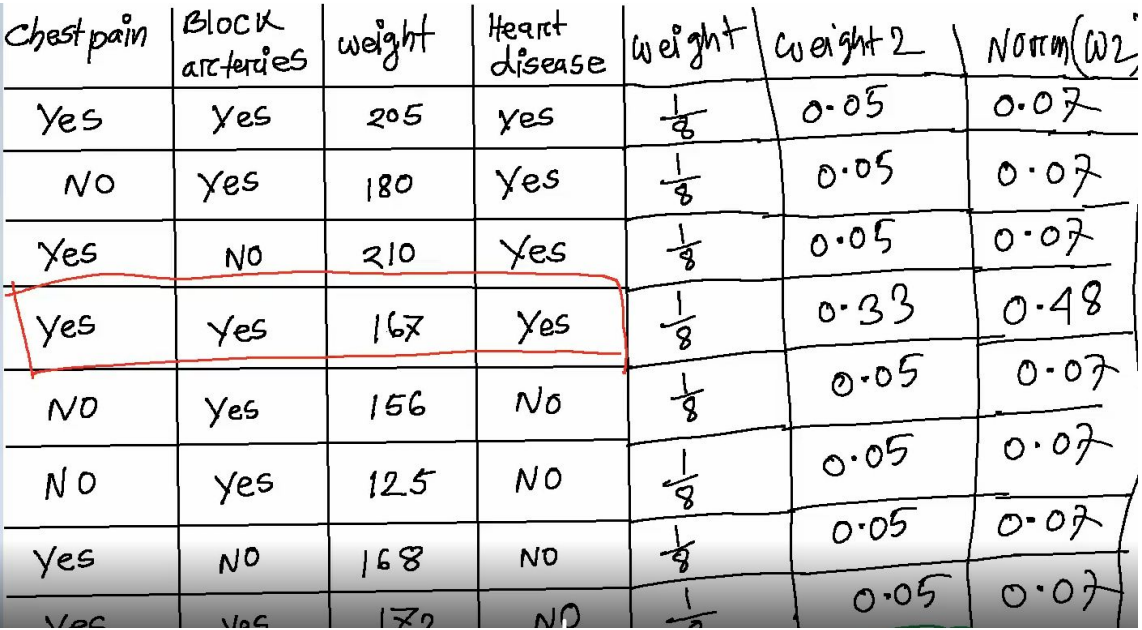
MisClassified : Weight will increase

Correct: Decrease

Now, Insert Updated Weight to the Table

Chest pain	Block arteries	weight	Heart disease	weight	weight 2
Yes	Yes	205	yes	$\frac{1}{8}$	0.05
NO	Yes	180	Yes	$\frac{1}{8}$	0.05
Yes	NO	210	Yes	$\frac{1}{8}$	0.05
Yes	Yes	167	Yes	$\frac{1}{8}$	0.33
NO	Yes	156	No	$\frac{1}{8}$	0.05
NO	Yes	125	NO	$\frac{1}{8}$	0.05
Yes	NO	168	NO	$\frac{1}{8}$	0.05
					0.05

Now Calculate Norm (Divide every weight by summation)

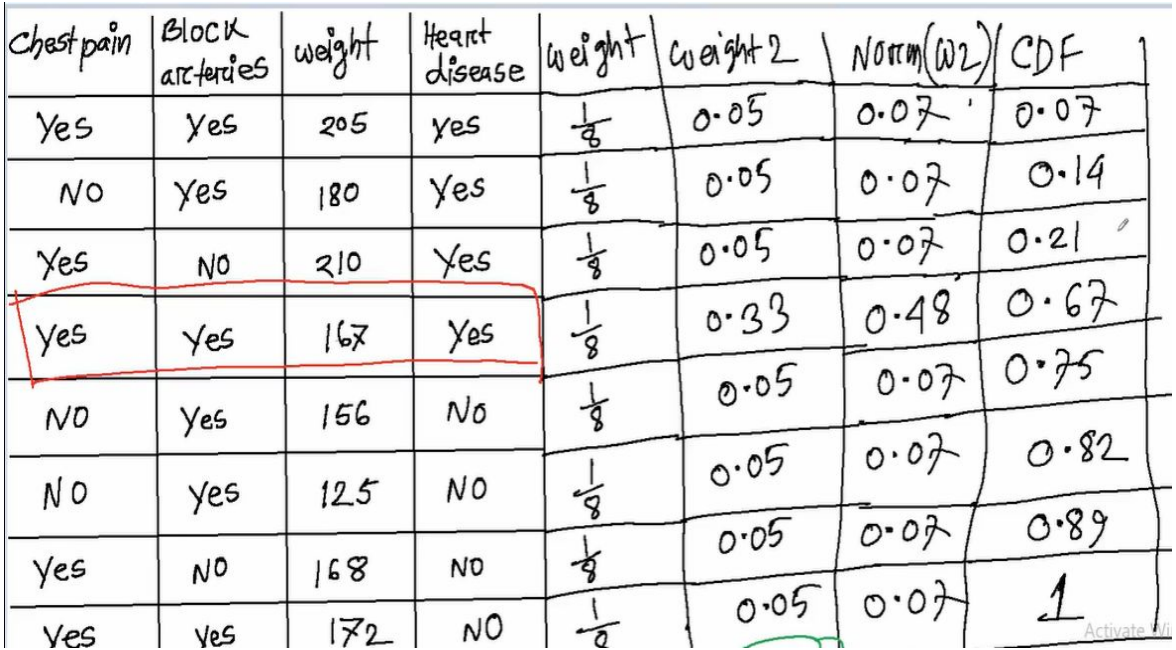


Chest pain	Block arteries	weight	Heart disease	weight	weight 2	Norm(w2
Yes	Yes	205	yes	1/8	0.05	0.07
NO	Yes	180	Yes	1/8	0.05	0.07
Yes	NO	210	Yes	1/8	0.05	0.07
Yes	Yes	167	Yes	1/8	0.33	0.48
NO	Yes	156	No	1/8	0.05	0.07
NO	Yes	125	NO	1/8	0.05	0.07
Yes	NO	168	NO	1/8	0.05	0.07
Yes	Yes	172	NO	1/8	0.05	0.07

Why Norm?

Because, Total Summation is not zero!

Now, Cumulative Distribution Function



Chest pain	Block arteries	weight	Heart disease	weight	weight 2	Norm(W2)	CDF
Yes	Yes	205	yes	1/8	0.05	0.07	0.07
NO	Yes	180	Yes	1/8	0.05	0.07	0.14
Yes	NO	210	Yes	1/8	0.05	0.07	0.21
Yes	Yes	167	Yes	1/8	0.33	0.48	0.67
NO	Yes	156	No	1/8	0.05	0.07	0.75
NO	Yes	125	NO	1/8	0.05	0.07	0.82
Yes	NO	168	NO	1/8	0.05	0.07	0.89
Yes	Yes	172	NO	1/8	0.05	0.07	1

Why CDF?

CDF is just a tool to decide who gets more attention.

Let's See the ***Range***

Chest pain	Block arteries	weight	Heart disease	weight	weight 2	Norm(W2)	CDF	Range
Yes	Yes	205	yes	$\frac{1}{8}$	0.05	0.07	0.07	0 - 7
NO	Yes	180	Yes	$\frac{1}{8}$	0.05	0.07	0.14	8 - 14
Yes	NO	210	Yes	$\frac{1}{8}$	0.05	0.07	0.21	15 - 21
Yes	Yes	167	Yes	$\frac{1}{8}$	0.33	0.48	0.67	22 - 67
NO	Yes	156	No	$\frac{1}{8}$	0.05	0.07	0.75	67 - 75
NO	Yes	125	NO	$\frac{1}{8}$	0.05	0.07	0.82	76 - 82
Yes	NO	168	NO	$\frac{1}{8}$	0.05	0.07	0.89	83 - 89
					0.05	0.07	1	90 - 100

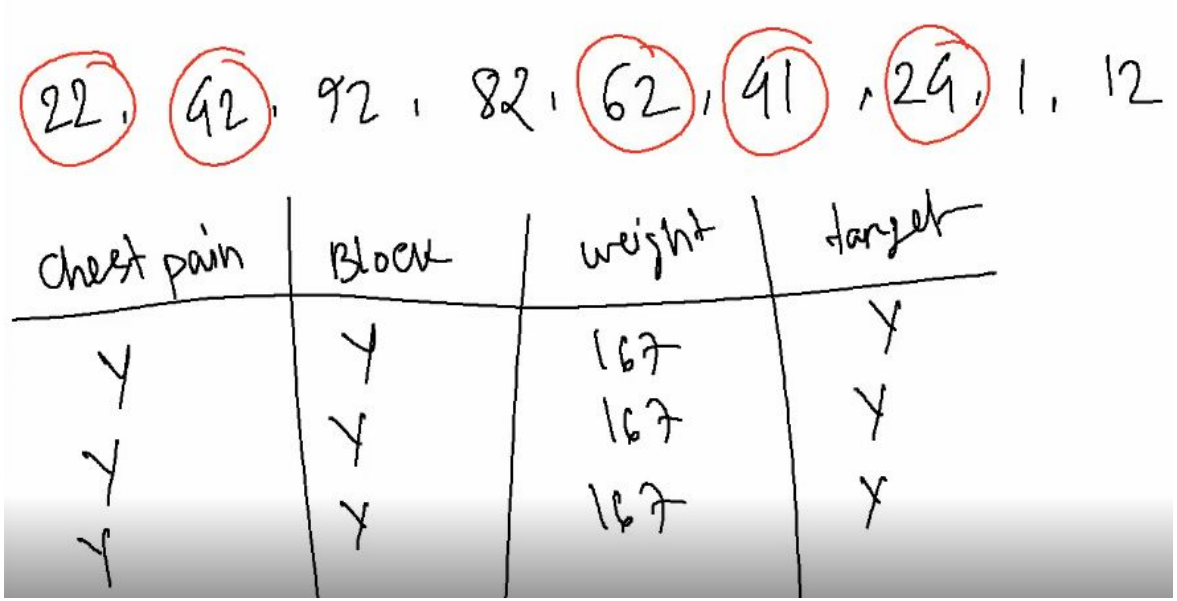
Now Backtrack and observe why Range is bigger

Chest pain	Block arteries	weight	Heart disease	weight	weight 2	Norm(W2)	CDF	Range
Yes	Yes	205	yes	1/8	0.05	0.07	0.07	0 – 7
NO	Yes	180	Yes	1/8	0.05	0.07	0.14	8 – 14
Yes	NO	210	Yes	1/8	0.05	0.07	0.21	15 – 21
Yes	Yes	167	Yes	1/8	0.33	0.48	0.67	22 – 67
NO	Yes	156	No	1/8	0.05	0.07	0.75	67 – 75
NO	Yes	125	NO	1/8	0.05	0.07	0.82	76 – 82
Yes	NO	168	NO	1/8	0.05	0.07	0.89	83 – 89
Yes	Yes	172	NO	1/8	0.05	0.07	1	90 – 100

Intuition Behind Range

Let’s Guess Number 0 to 100 : (it should be random!)

Let’s Create a New Dataset now!



Now we'd force this stump to classify our misclassification!

To visualize it better