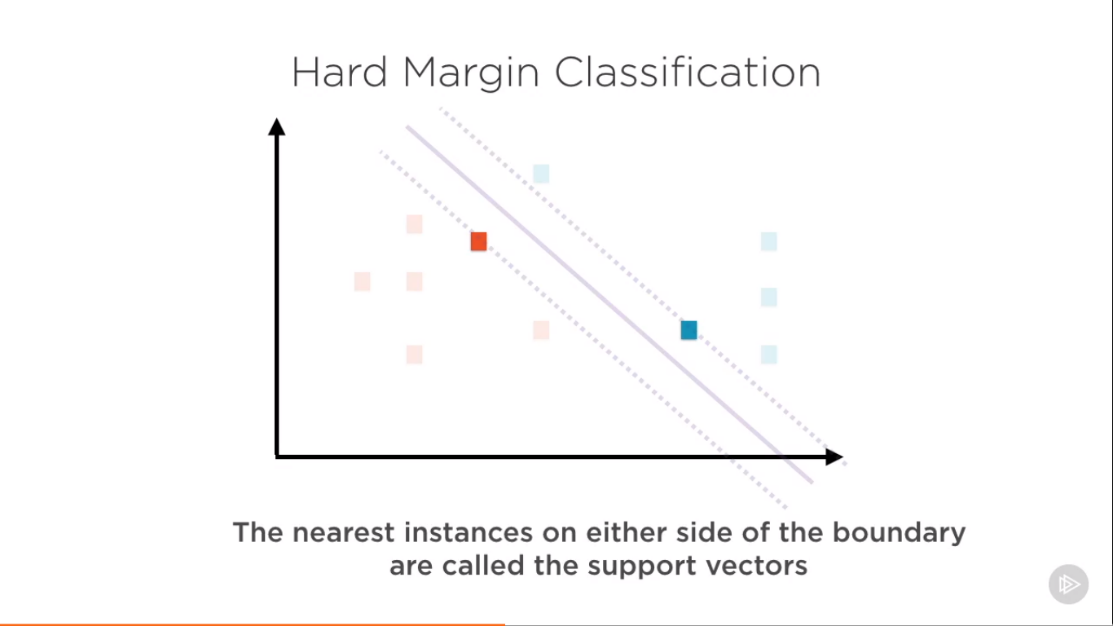
**SVM and Gradient Boosting Models**

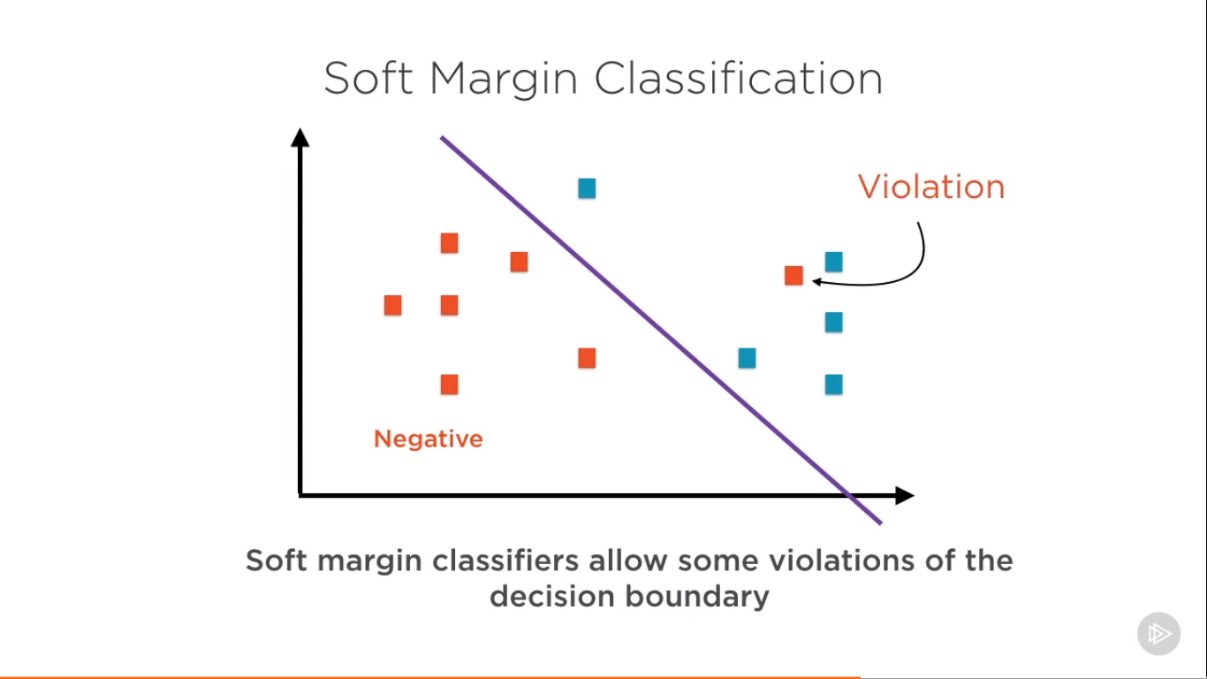
**SVM**

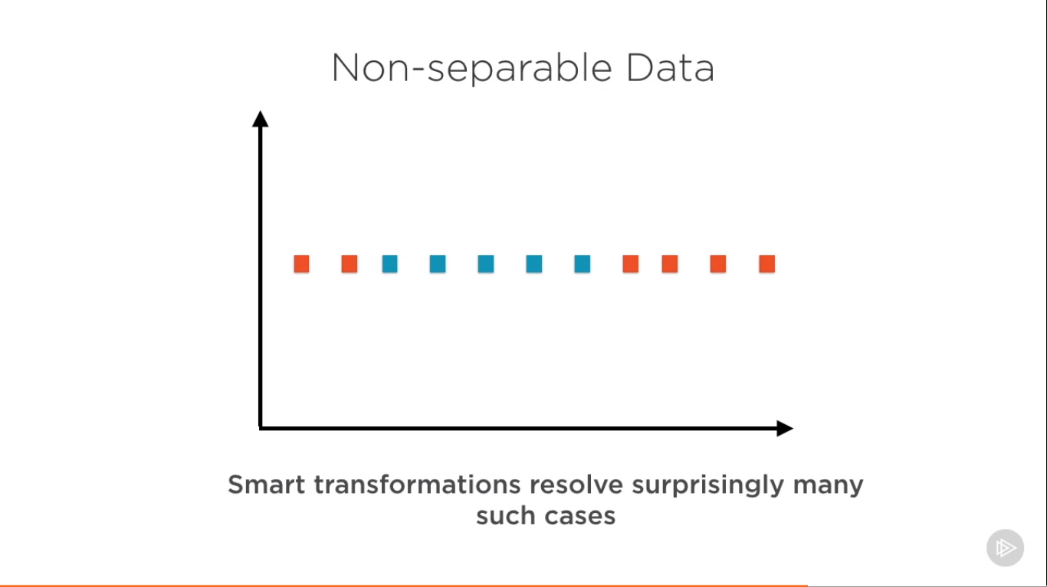
****

For good SVM we want the support vectors to be as far as possible. We want to fit the widest boundary.

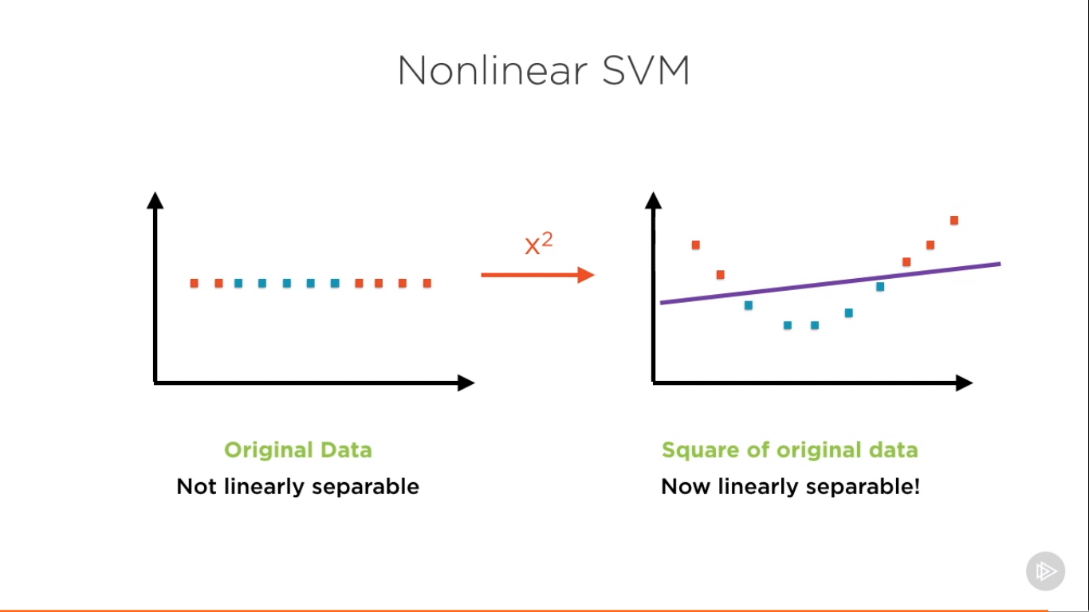
**Note** – Hard line classifier do not allow any of the outlier while soft margin classifier does.

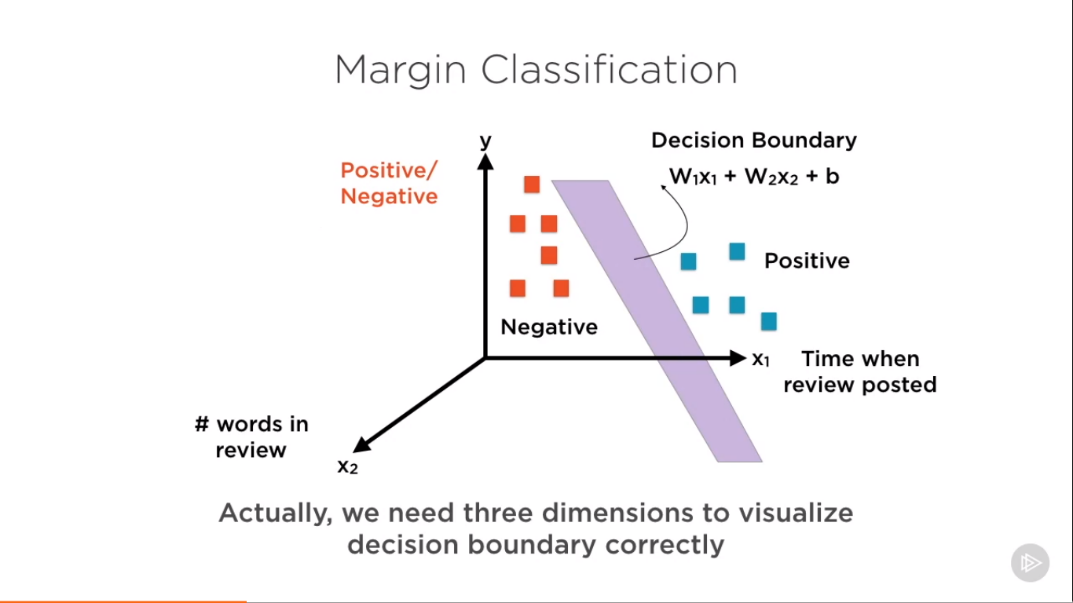
Hard classifier requires perfectly linear separability in data.





SVM uses **kernel trick** to extend to almost any data.

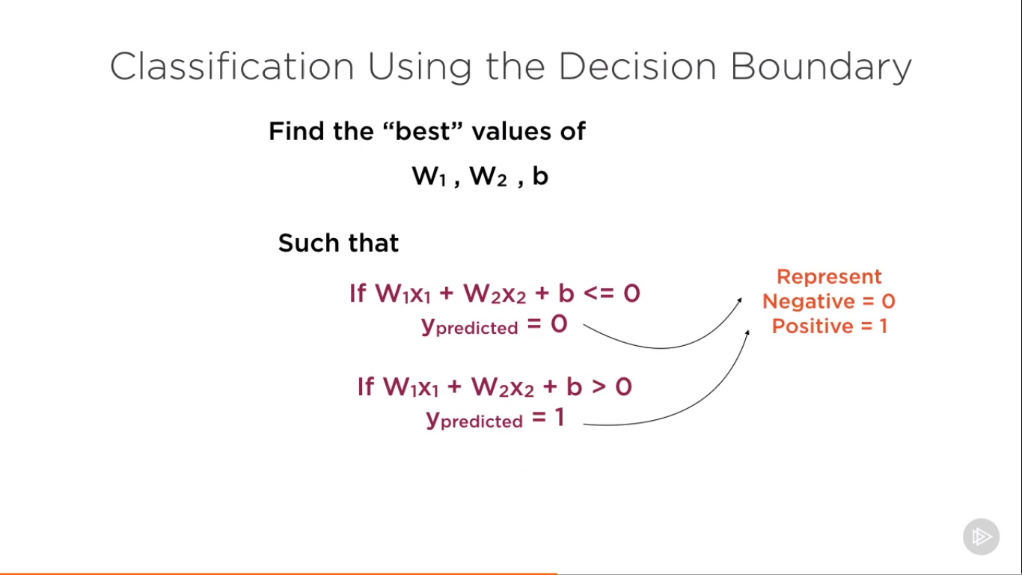


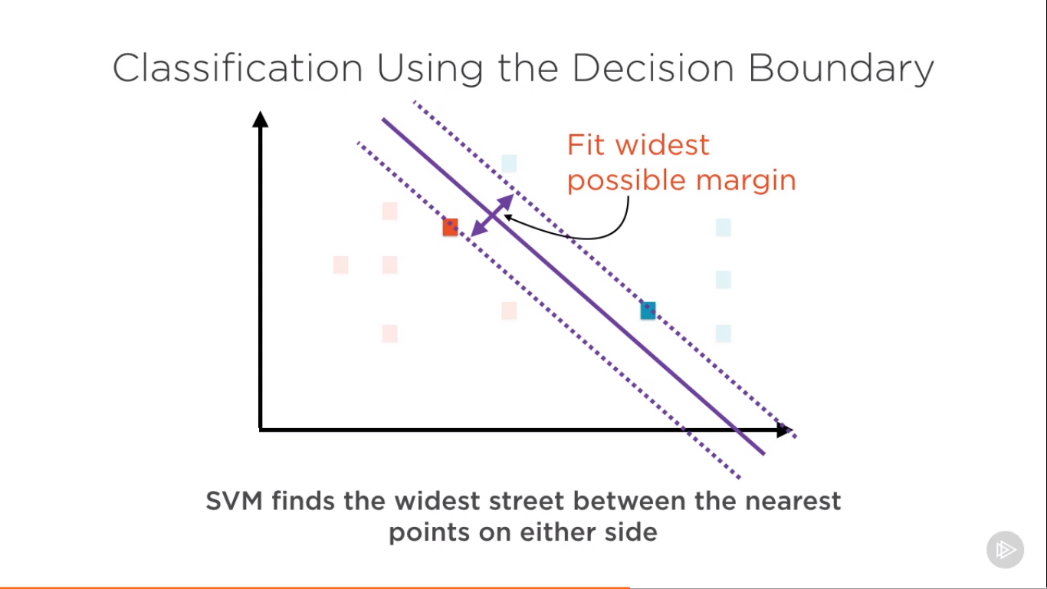


**If w1x1 + w2x2 + b > 0 , positive reviews**

**if w1x1 + w2x2 + b <= 0, negative reviews**

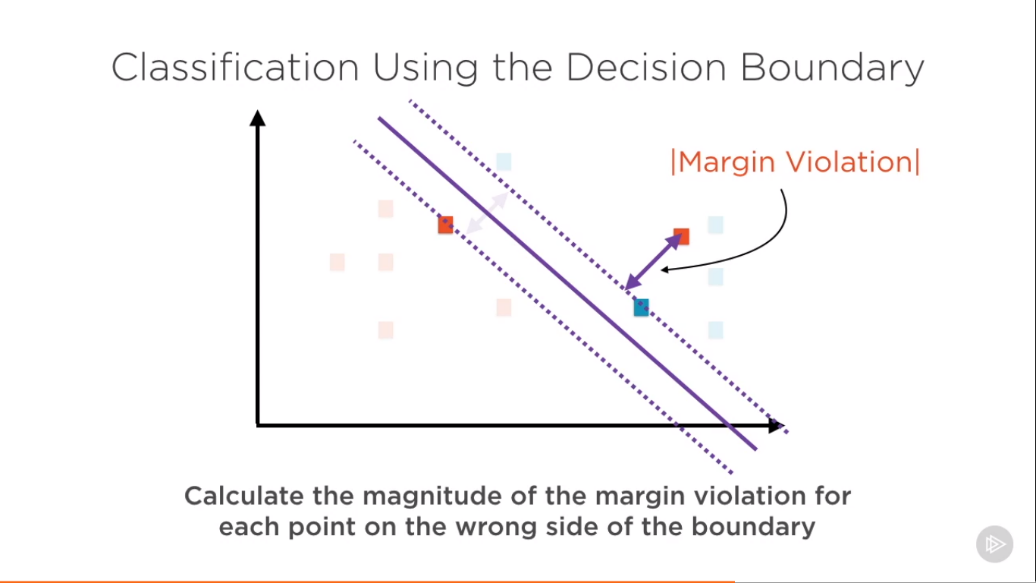
Objective of the SVM classifier is to find the **best** values of w1, w2, b (**optimization problem**)





**“best”** – widest margin between support vector

But best must also avoid or minimize **outliers** ( by penalizing them during the optimization )



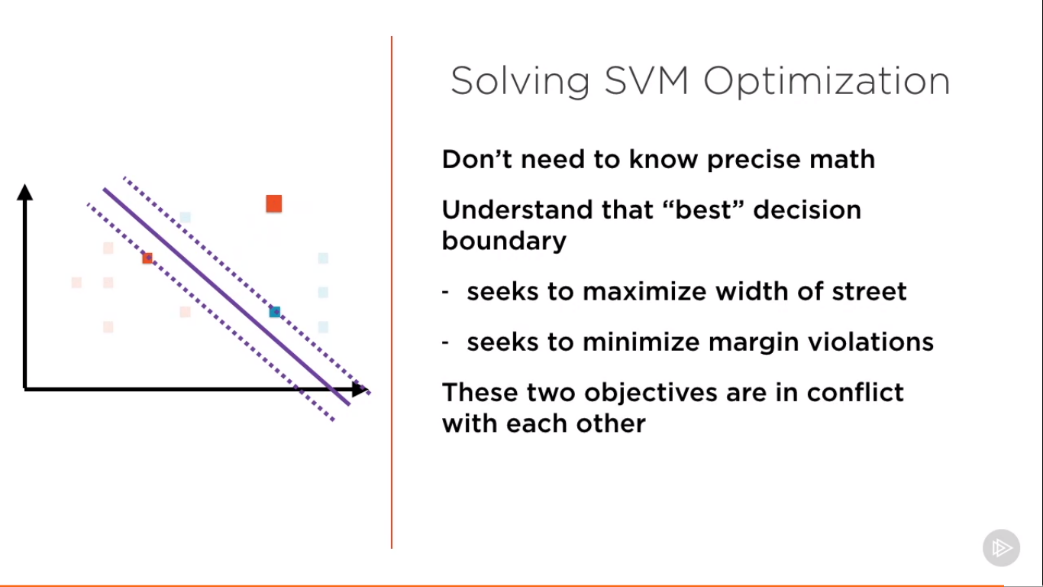
Calculate the magnitude of the margin violation for each point on the wrong side of the boundary.

Multiply this magnitude of margin of violation by penalty factor C.

**Penalty = C \* |Margin Violation|**

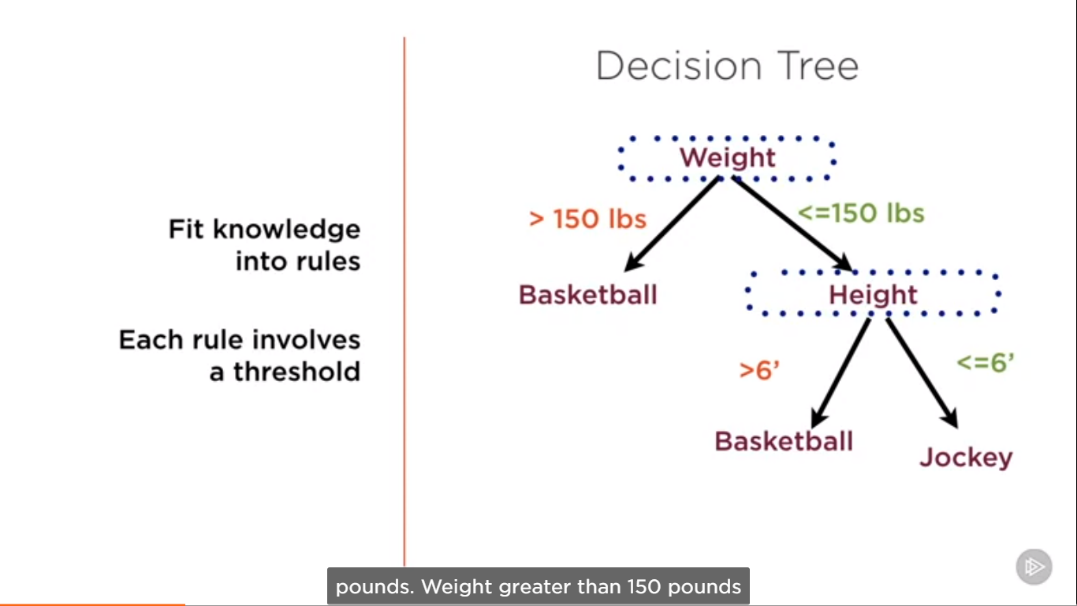
Very large value of C – Hard margin classifier

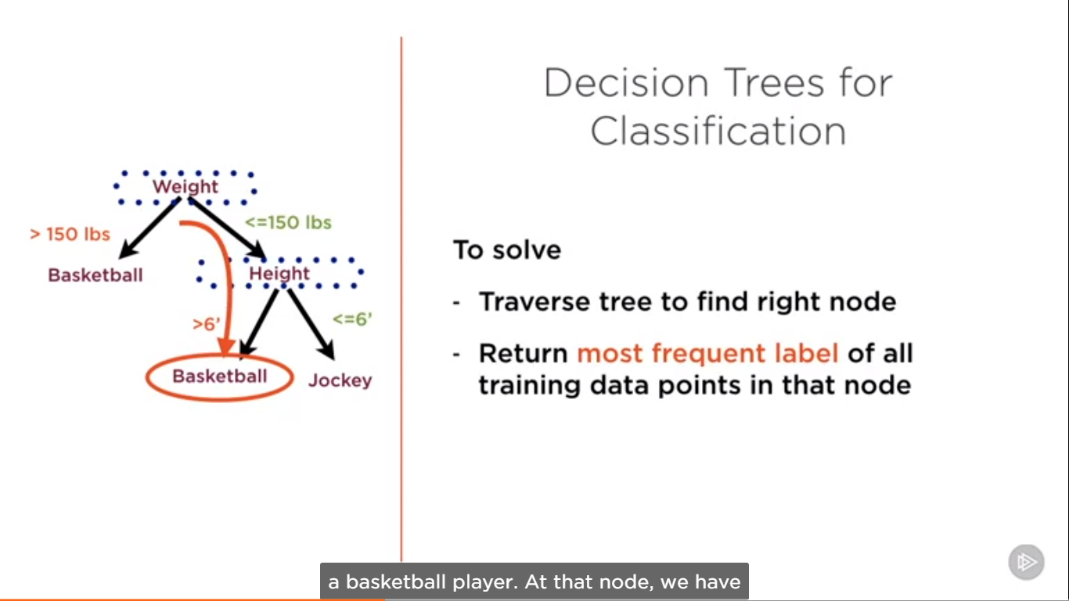
Very small value of C – Soft margin classifier

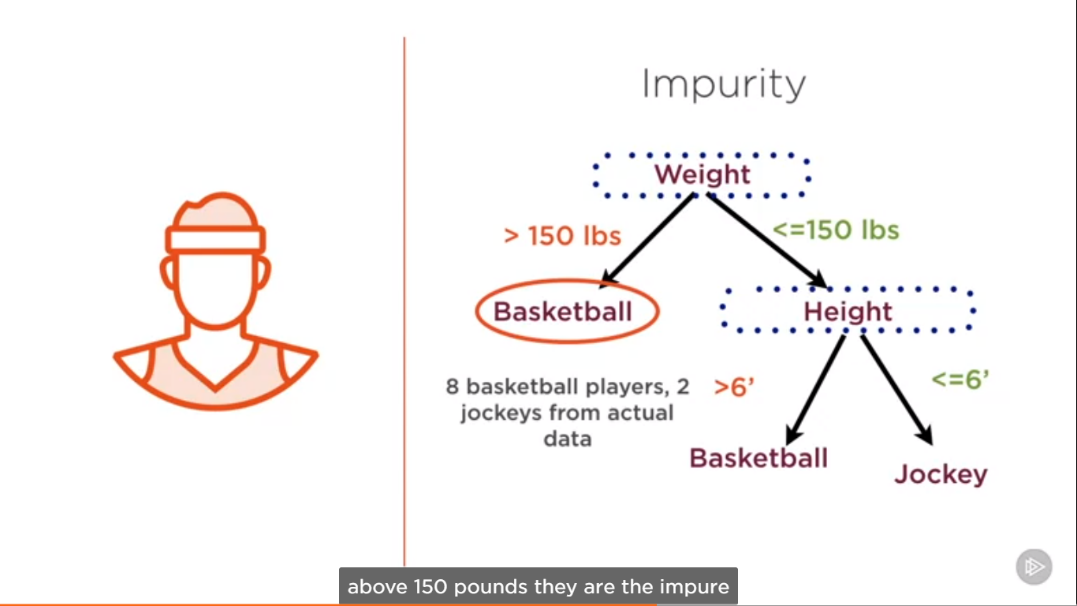


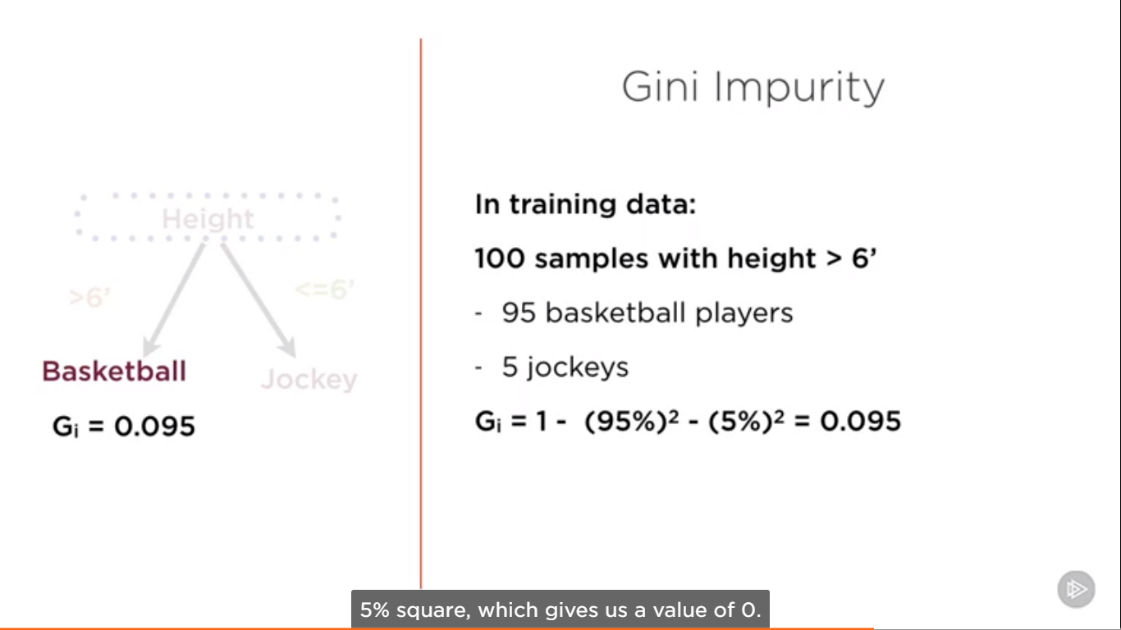
**Demo of SVM:**

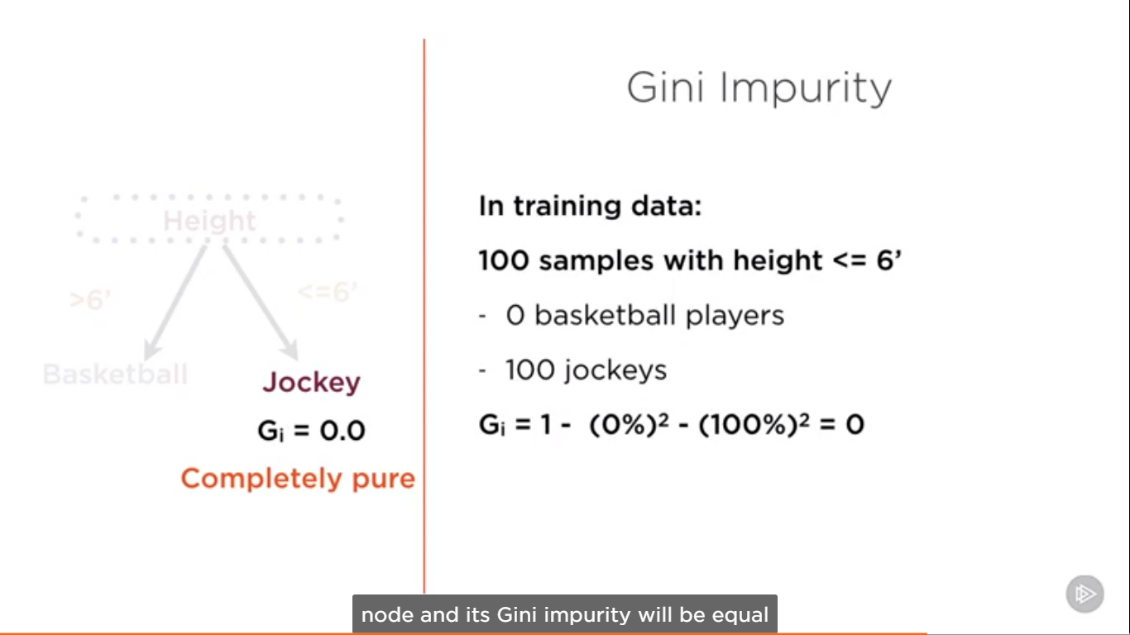
**Decision Trees**

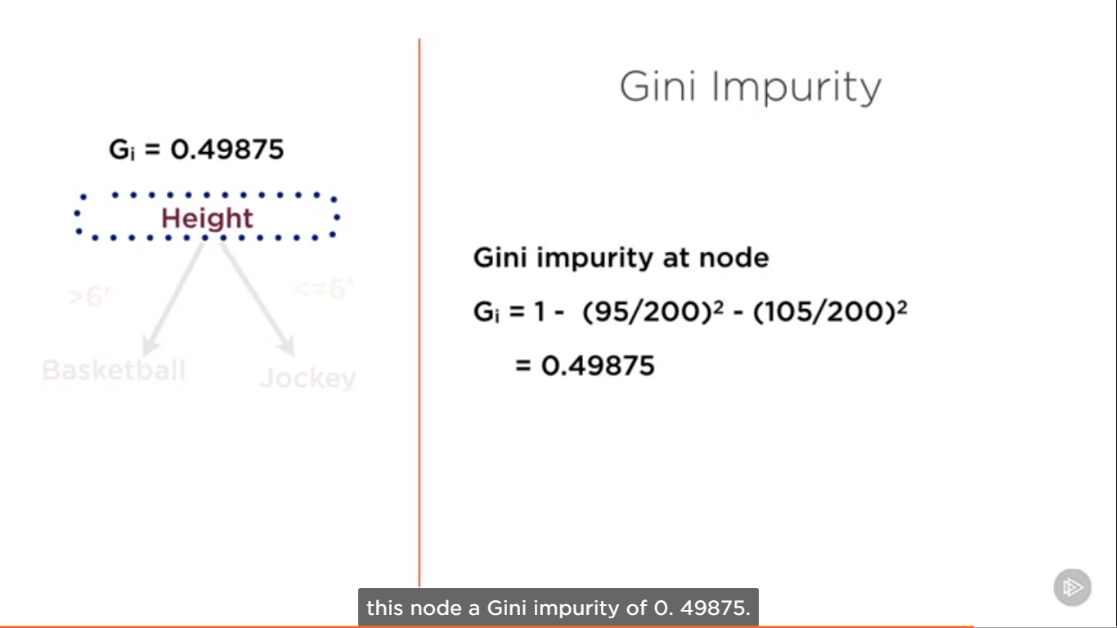
****

****

****

****

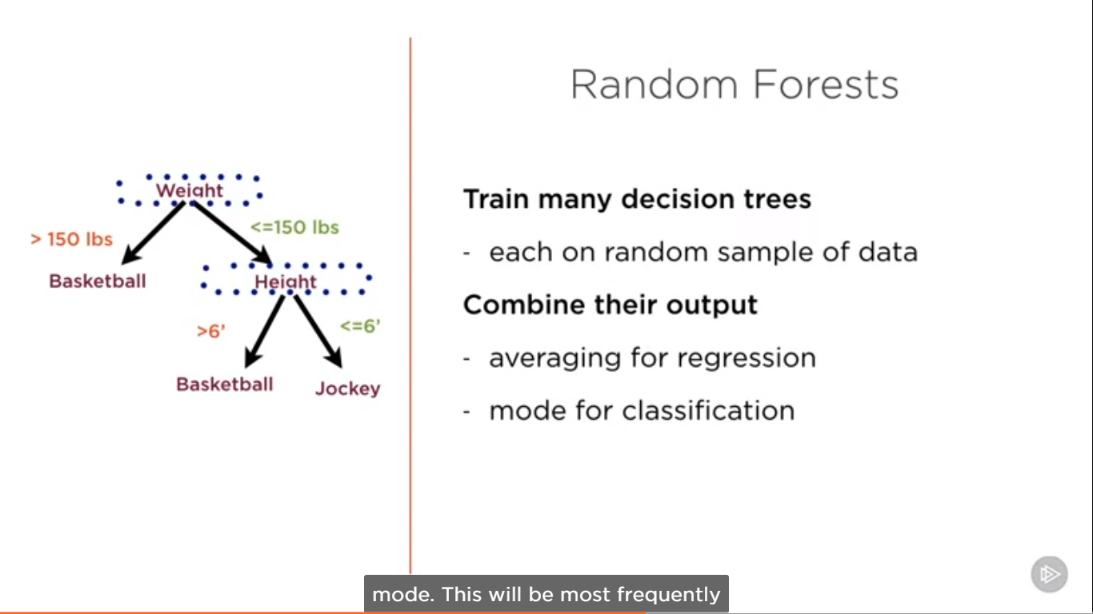
****

****

**Prone to overfitting ( common risk with non parametric )**

**Unstable – Small changes in data cause big changes in model**

**Random Forest**

****

Extremely powerful technique

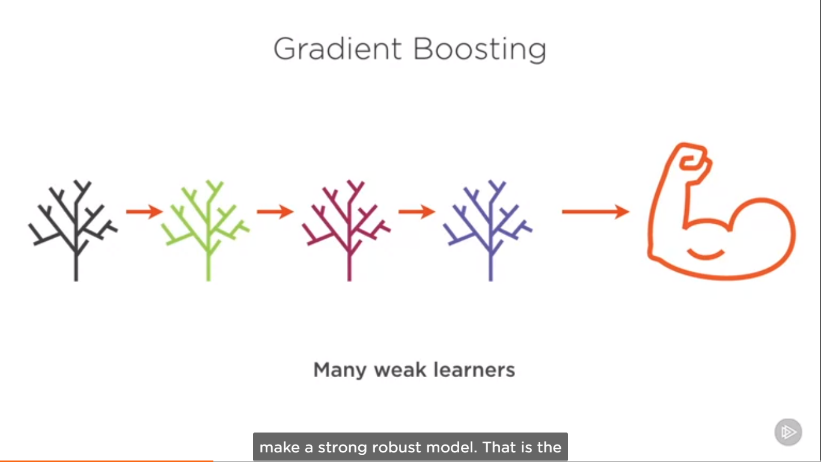
Example of **ensemble** learning

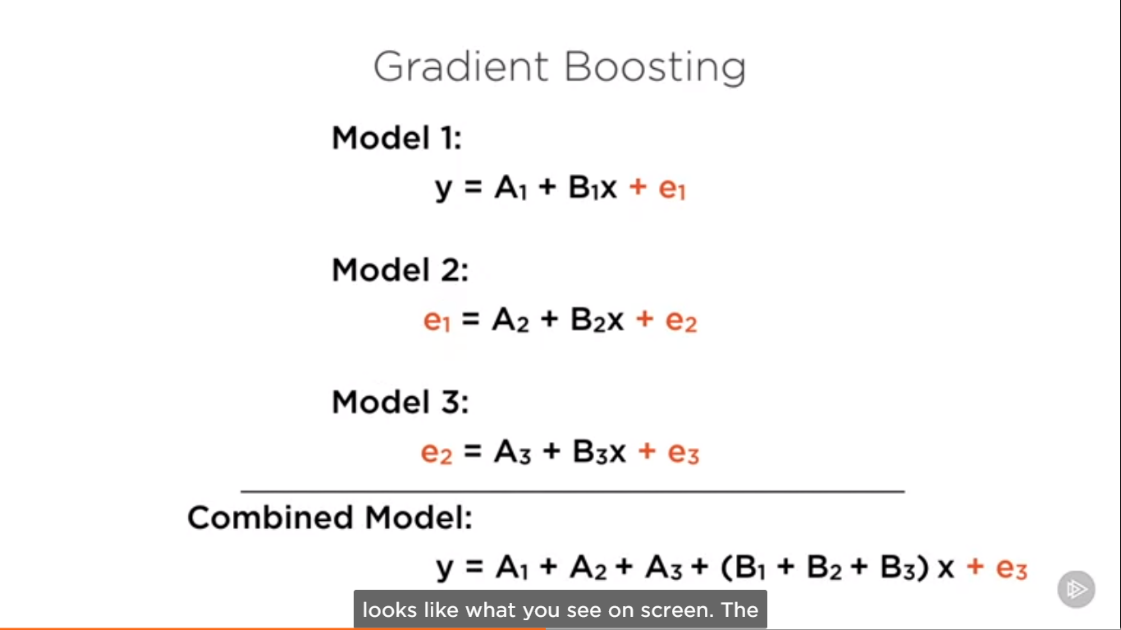
Individual trees should be as **different as possible**

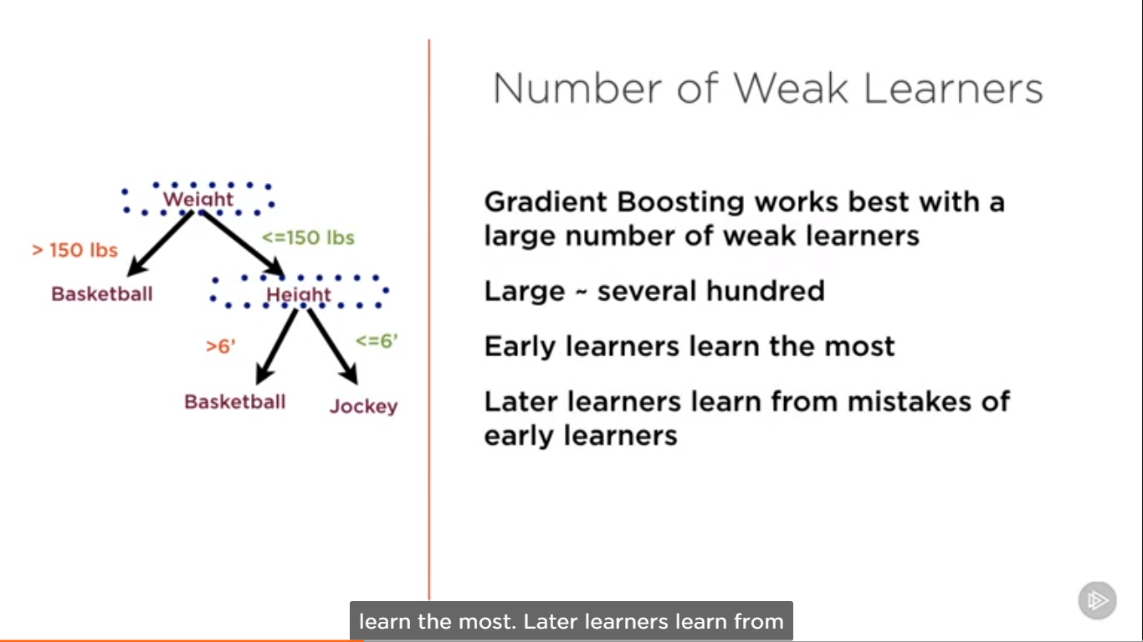
**Gradient Boosting**

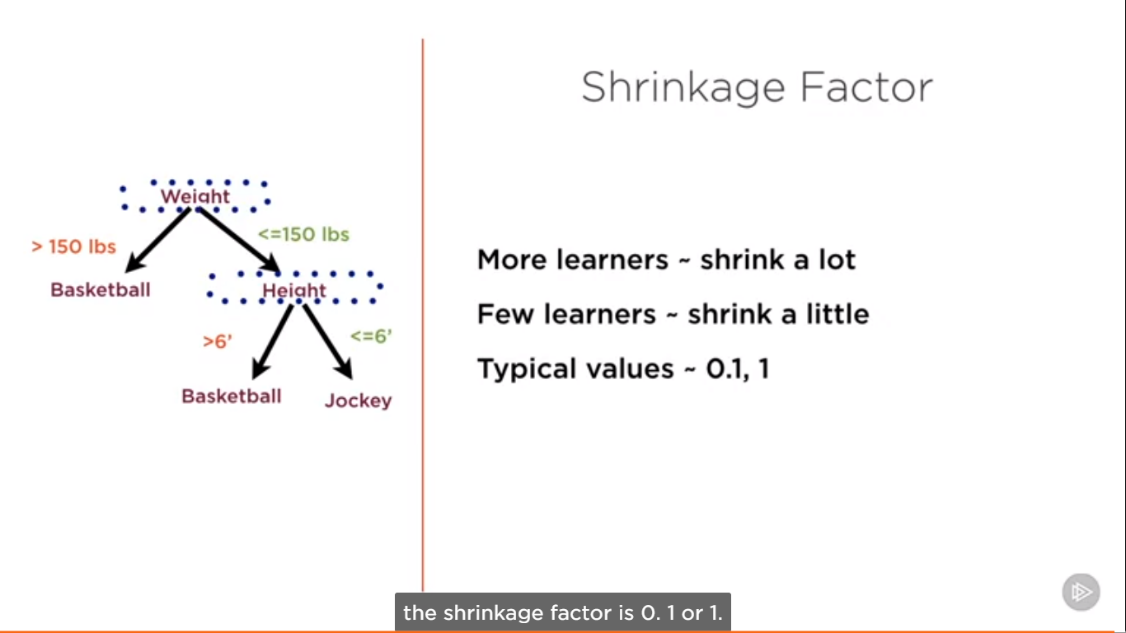
Build up your weaknesses until they become your strong points.

It uses many weak decision trees to make a strong one.









More Learners – shrink a lot

Few Learners – shrink a little

Typical value – 0.1, 1

**Demo: Gradient Boosting in scikit learn**