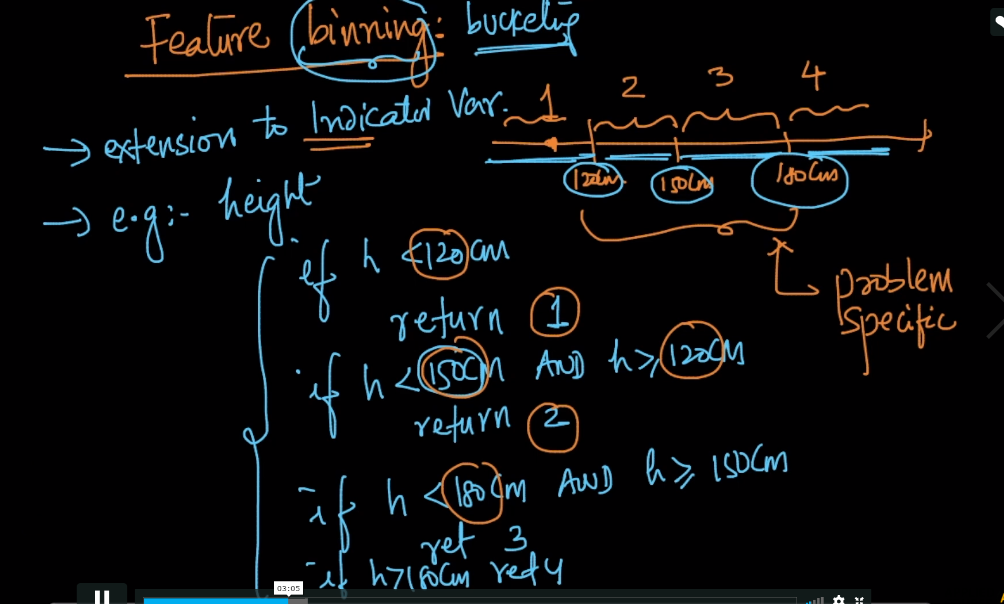
**Feature binning**

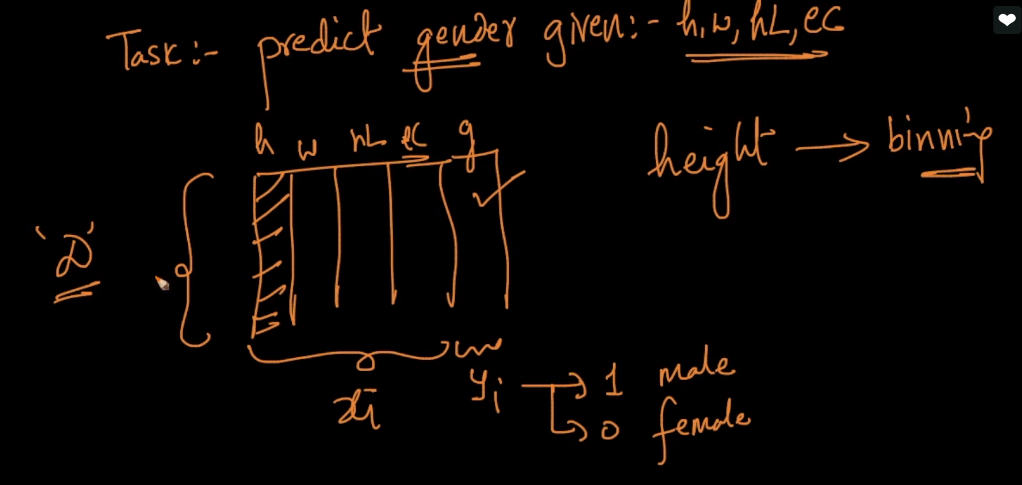
Feature binning is also known as bucketing.

Lets take a extension to indicator variable. We take a example of height and create many indicators.



So if we have a task to predict gender given features height, weight, hair length, eye color

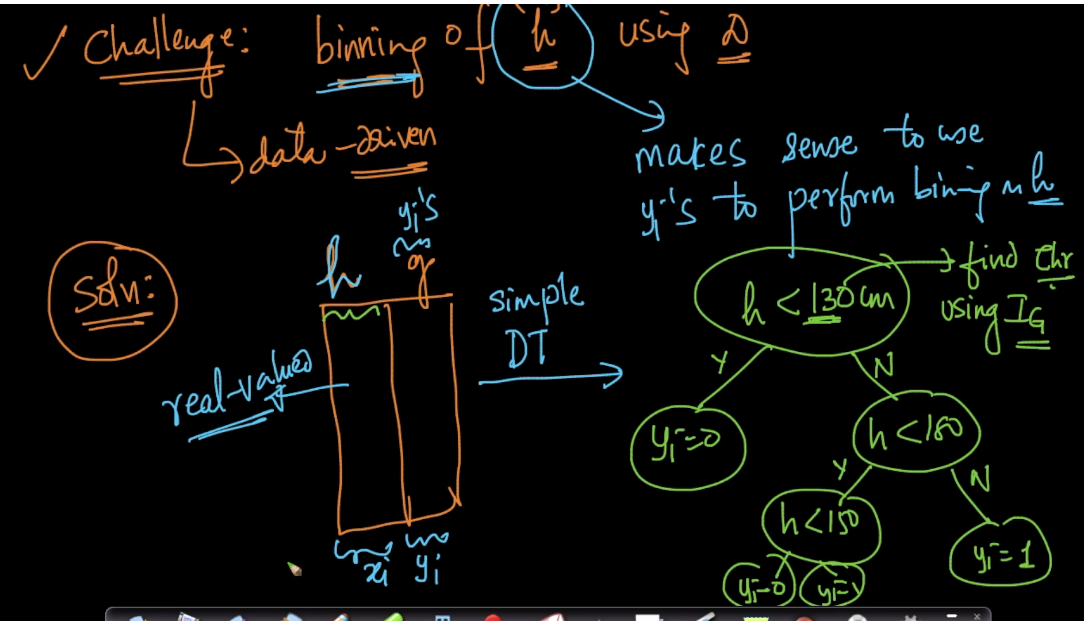
Now just focus on binning of height feature.



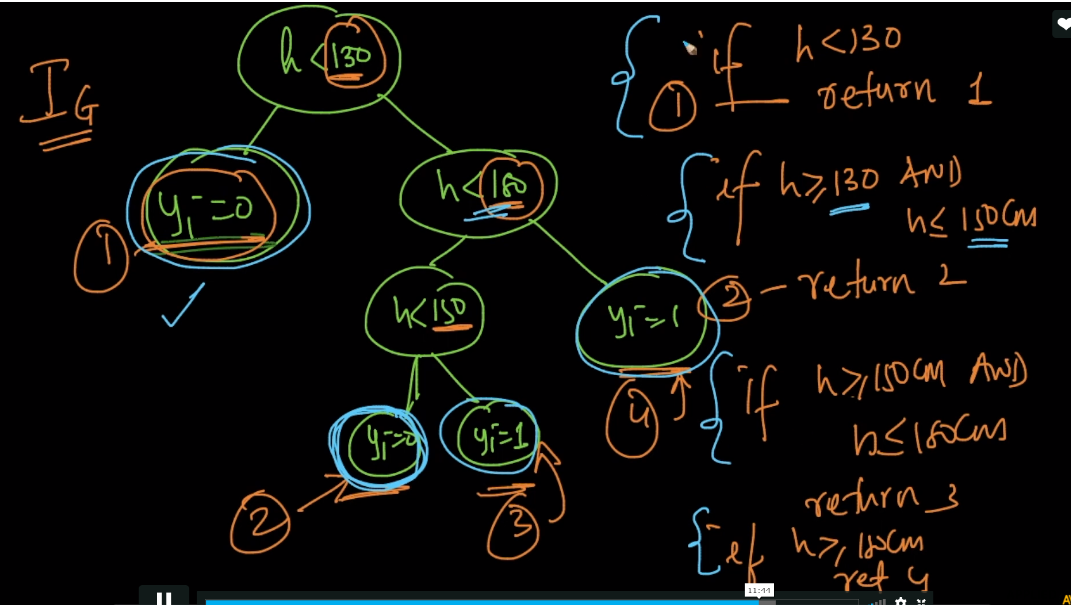
So it’s a challenge of binning of height h using D

So the solution of this challenge is to use feature h and yi’s (gender) to perform bining of h.

Now what we do is just apply simple decision tree on this

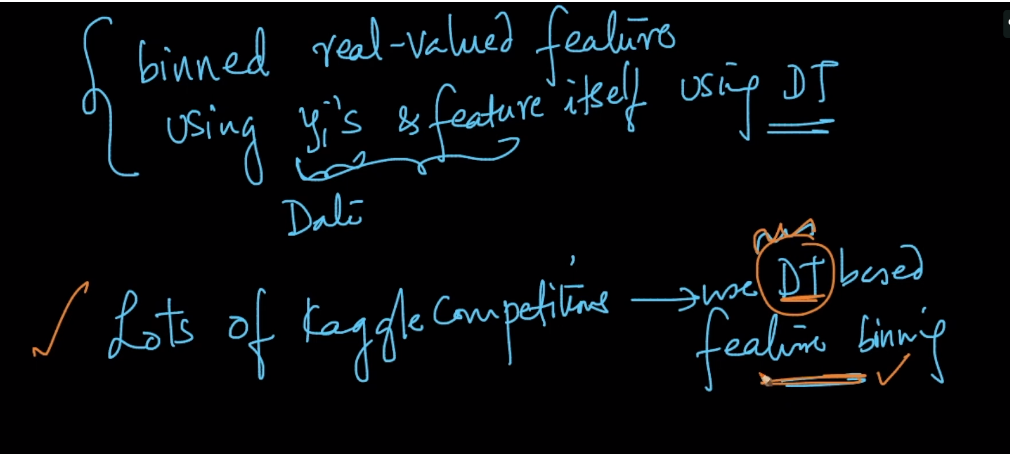


Now decision tree generate conditions based on threshold using IG and return us output as shown in below figure.



So what we just did is binned real-values features using yi’s and feature itself using DT.

So models like DT not only use for classification but it can also use for other purpose like feature binning and feature engineering



Comments :

