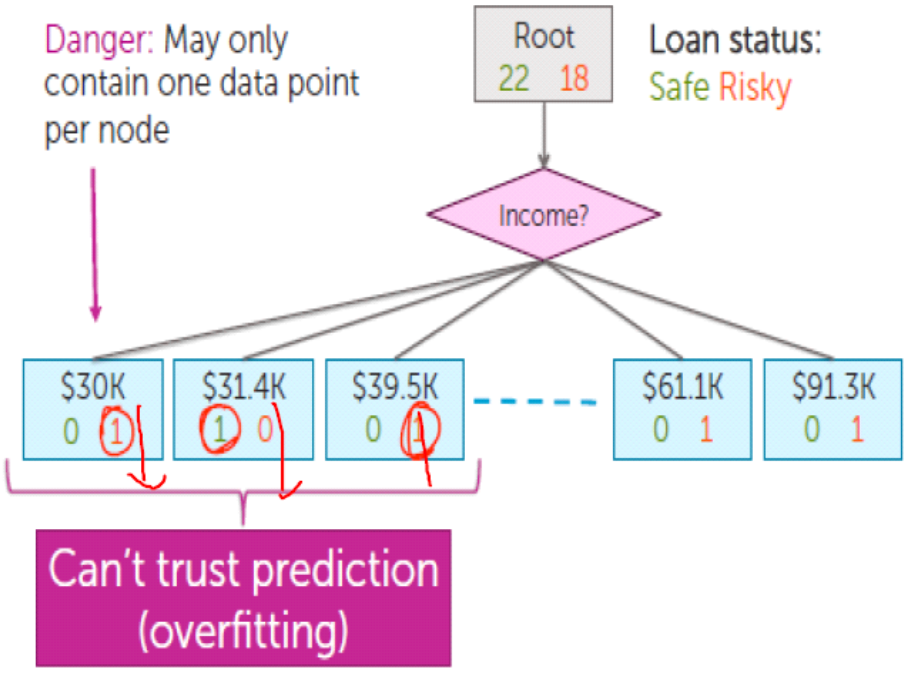
09 March 2018
12:32

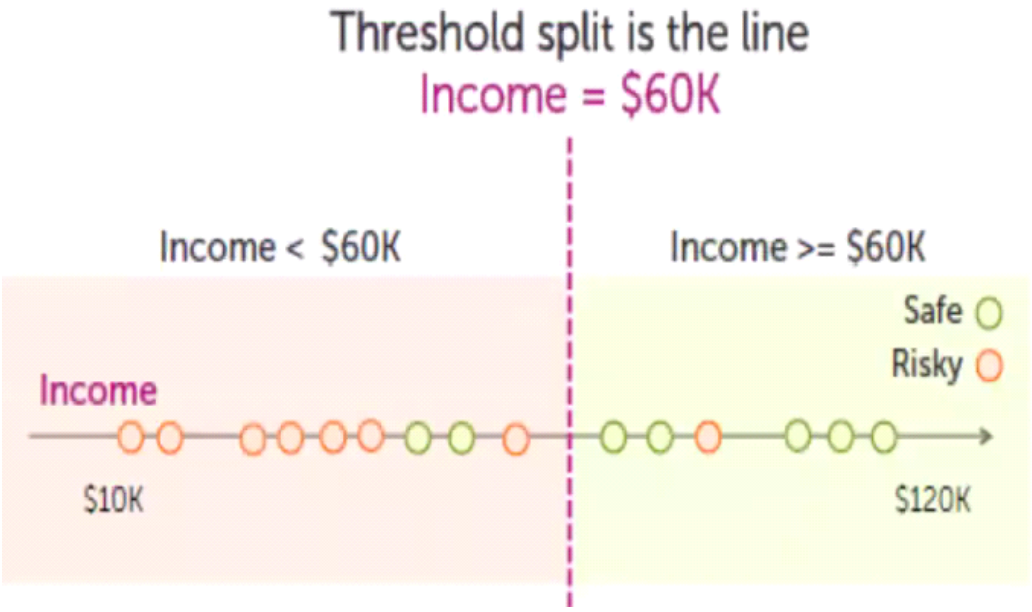
How to work if your independent variables are continuous?

How do we use real values inputs?

Income	Credit	Term	y
$105 K	excellent	3 yrs	Safe
$112 K	good	5 yrs	Risky
$73 K	fair	3 yrs	Safe
$69 K	excellent	5 yrs	Safe
$217 K	excellent	3 yrs	Risky
$120 K	good	5 yrs	Safe
$64 K	fair	3 yrs	Risky
$340 K	excellent	5 yrs	Safe
$60 K	good	3 yrs	Risky

Split on each numeric value?





Prediction values for above split is Risky (7) and Safe (5)

Miss classification error = $\frac{2+1}{40} = 0.075$,

Similarly you can calculate for <$50k and >=$50k, <$40k and >=$40k, <$70k and >=$70k

Split on the different range values such that where ever you get Minimum classification error. This is finalized.

ADVANTAGES:

- high interpretability - easy to expalin
- high visualization power
- close to resemblance to human-thinking process.
- can handle qualitative predictors without creating dummy variables.

DISADVANTAGES:

- lower prediction accuracy
- non-robust with small changes in data