DT

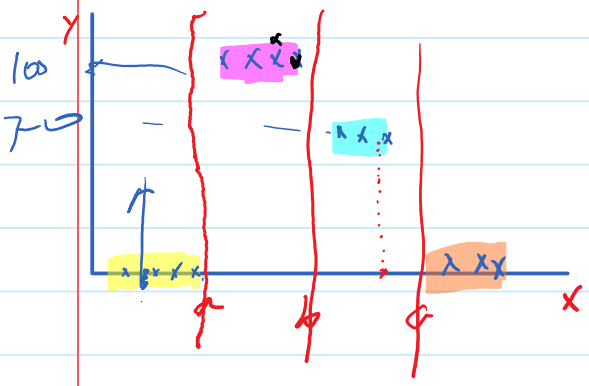
Sunday, September 20, 2020 6:14 PM

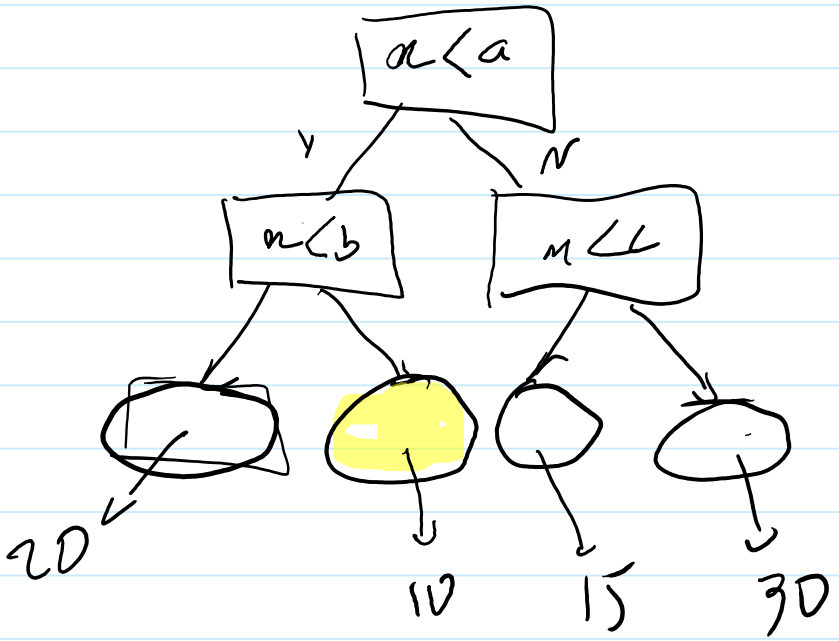
Decision Tree
- Random forest
- extreme randomizer tree
- adaboost random forest

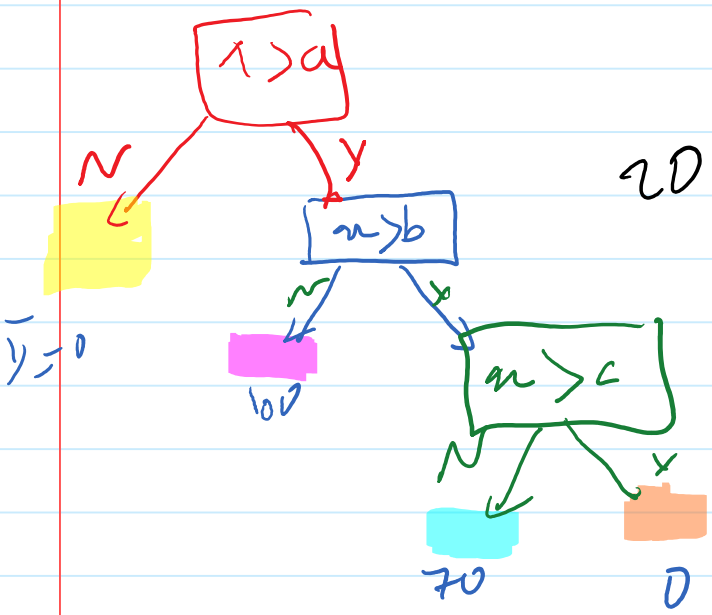
Xgboost
LGBM
Catboost

DecisionTreeRegressor

From <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>

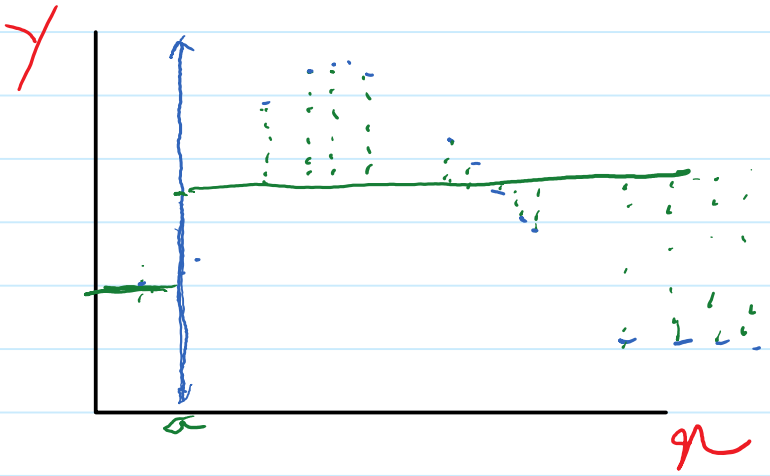






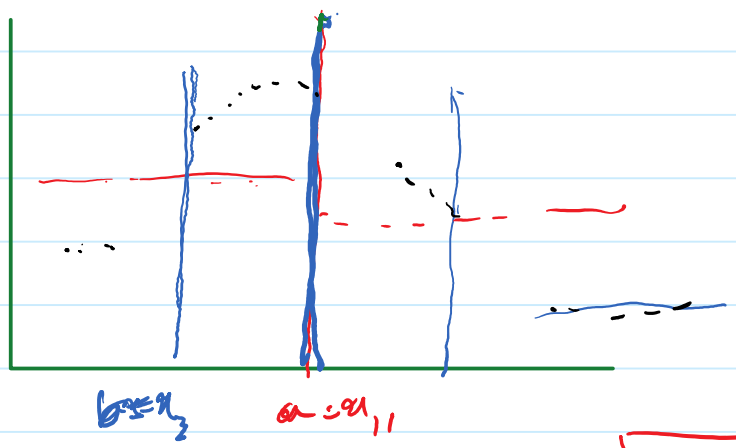
DT

Sunday, September 20, 2020 6:30 PM

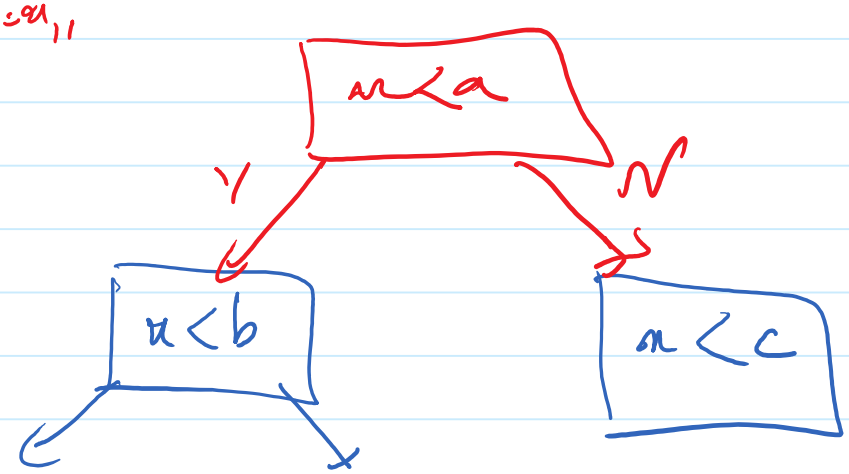


$x < a$

$$MSE = (Y_1 - \bar{Y}_{x<a})^2 + (Y_2 - \bar{Y}_{x<a})^2 + (Y_3 - \bar{Y}_{x<a})^2 + \dots$$



MSE = ✓

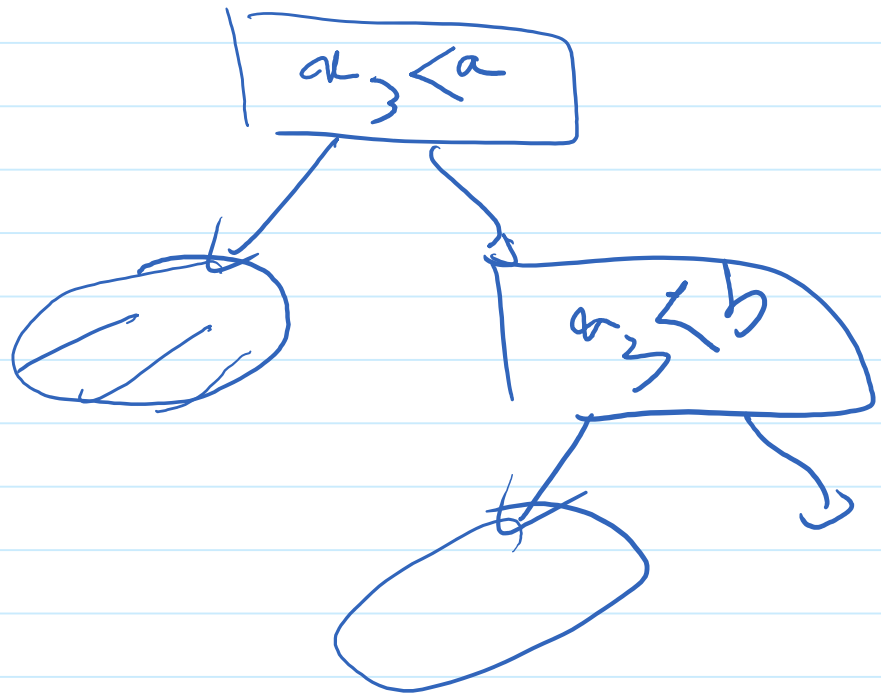


DT

Sunday, September 20, 2020 6:38 PM

x_1 x_2 x_3 .. x_{10} y

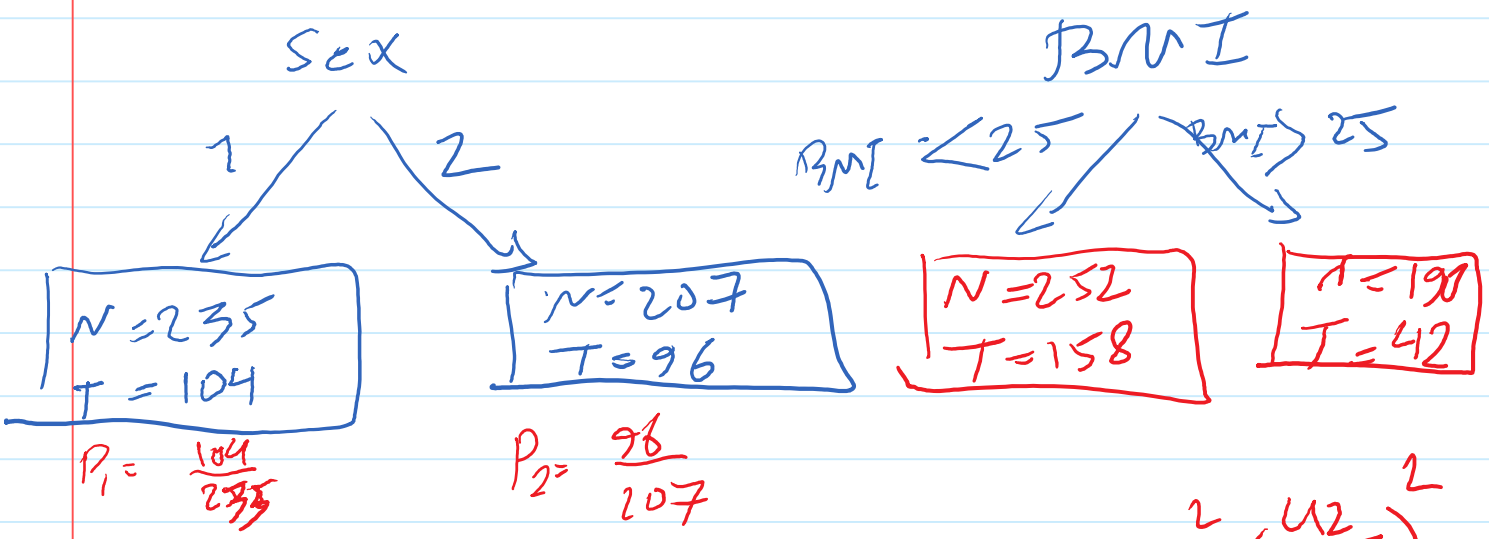
(5) x_3 (3) x_{10} (1) x_1



DT_clf_example_diabetes

Sunday, September 27, 2020 6:10 PM

$N = 443$



$P_1 = \frac{104}{235}$ $P_2 = \frac{96}{207}$

Gini

$1 - P_1^2 - P_2^2$

$1 - \left(\frac{104}{200}\right)^2 - \left(\frac{96}{200}\right)^2$

0.2704 0.4992

Gini = 0.59

$1 - \left(\frac{158}{200}\right)^2 - \left(\frac{42}{200}\right)^2$

0.6241 − 0.0441

0.5799

Gini = 0.56

$f(x,y) = 1 - x^2 - y^2$

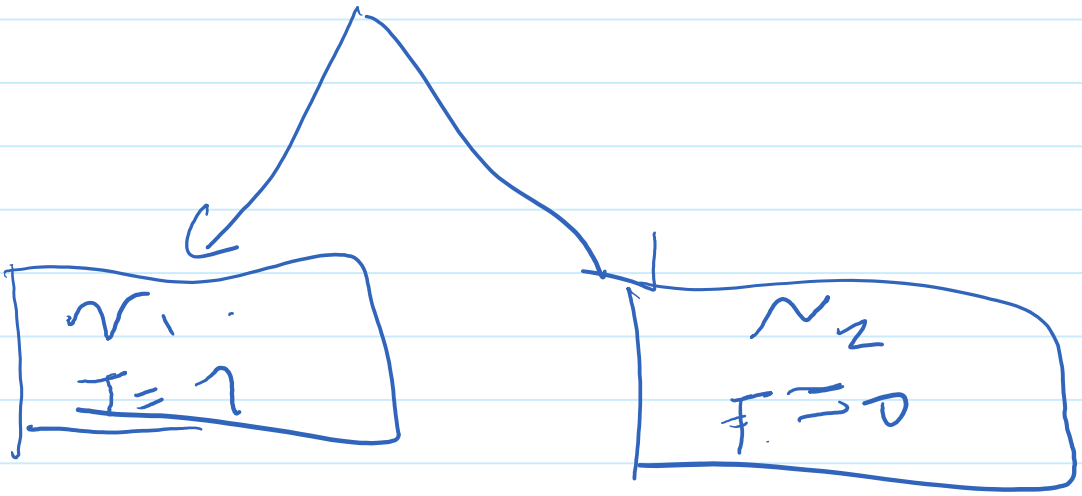
$x + y = 1$

$x = y = 0.5$

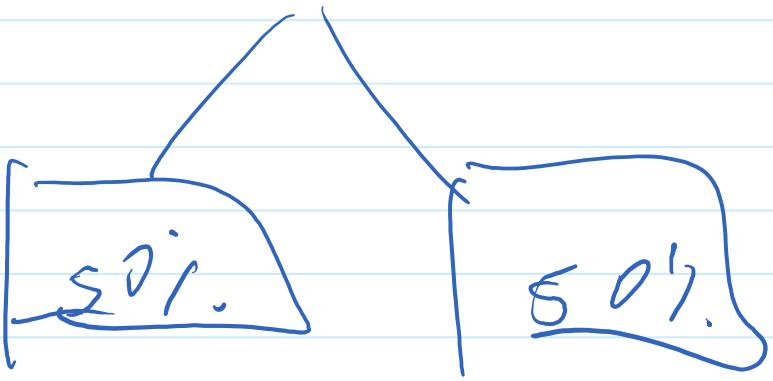
$\min f(x,y) = \text{or} \begin{matrix} x=1 \\ y=1 \end{matrix}$

DT_clf_example_diabetes

Sunday, September 27, 2020 6:32 PM



Gini = 0



DT_clf_example_diabetes

Sunday, September 27, 2020 6:44 PM