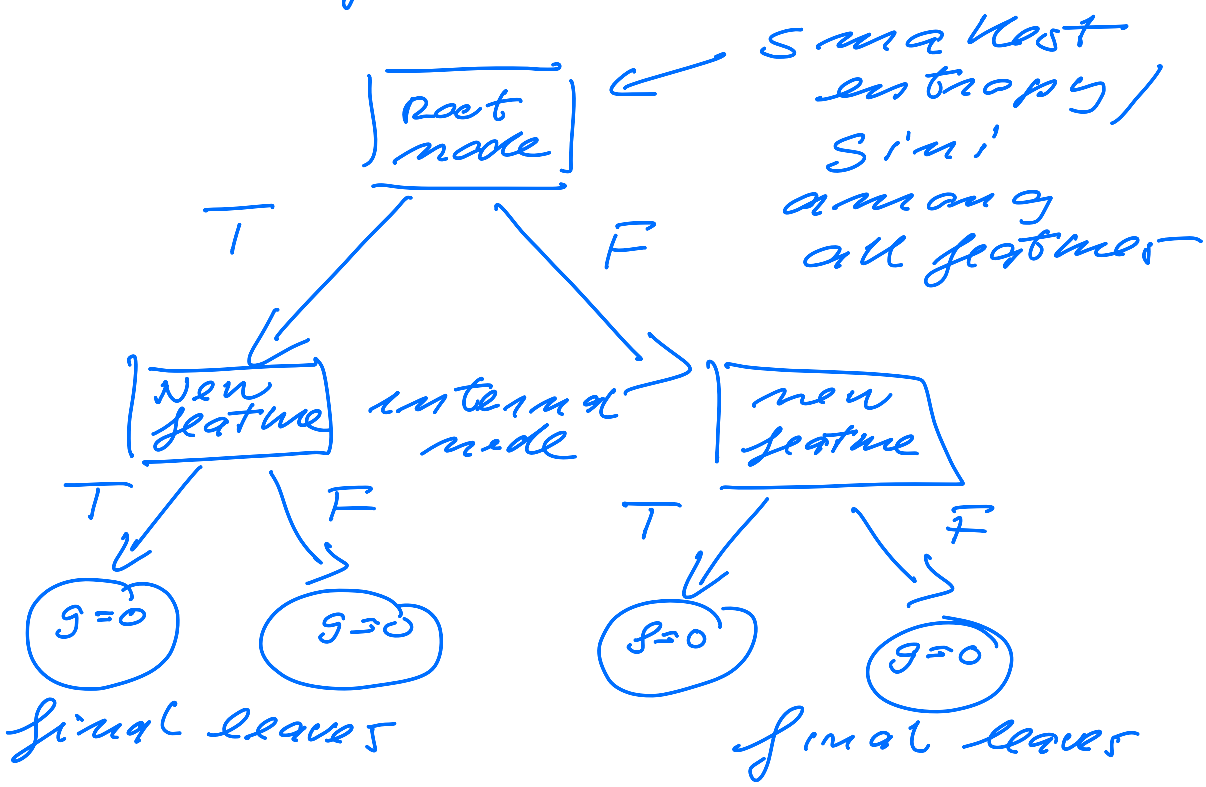
FYS-STK4155, Nov 4, 2022

Classification & Regression Tree (CART)

Classification



$$p(\text{outlook} = \text{sunny}) = 5/14$$

$$p(\text{outlook} = \text{rainy}) = 5/14$$

$$p(\text{outlook} = \text{overcast}) = 4/14$$

if (sunny & Ride) = 3/5

if (sunny & no) = 2/5

$$g = 1 - (3/5)^2 - (2/5)^2 = 12/25$$

if (overcast & Ride) = 4/4

if (overcast & no) = 0

$$g = 1 - 1 - 0 = 0$$

if (Rain & ride) P = 3/5

if (Rain & no) P = 2/5

gini index = 12/25

gini for outlook feature

$$5/14 \cdot 12/25 + 5/14 \, \frac{12}{25} + \frac{4}{14} \times 0$$

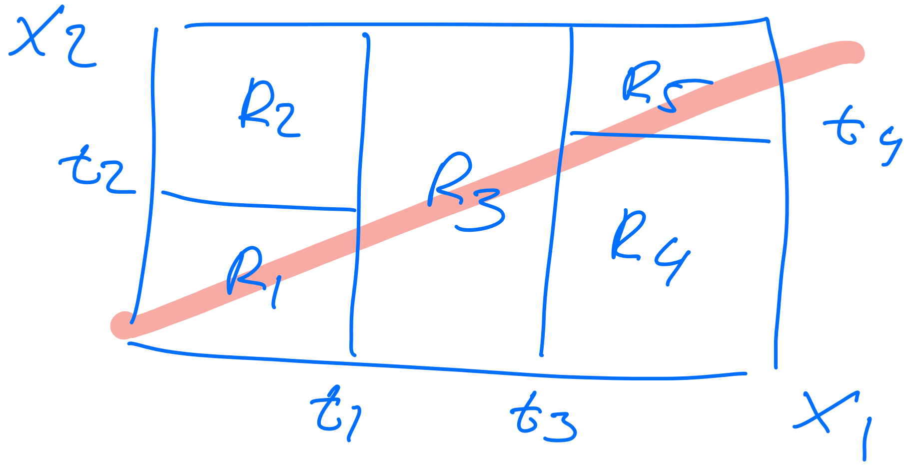
$$= 0,343$$

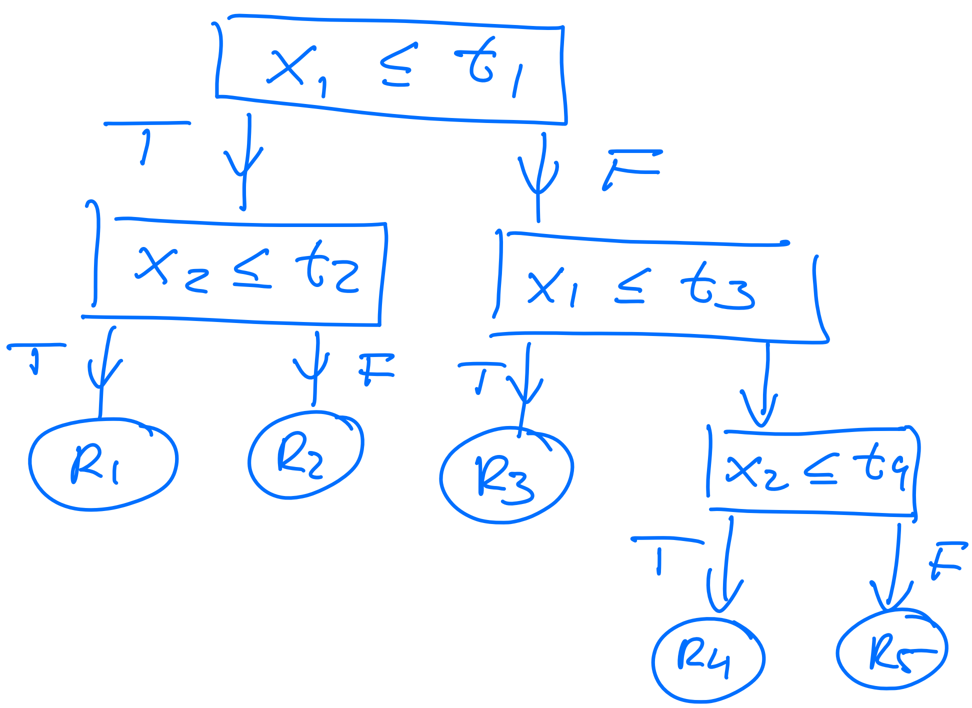
Humidity = 0.672

wind = 0.522

Temp = 0.405

Regression tree





Boosting

Basic philosophy is to use a simple function/approximation and then iterate in order to improve the reproduction of the data

— Regression case:

$$MSE = \frac{1}{n} \sum_i (y_i - \tilde{y}_i)^2$$

$$= \frac{1}{n} \sum_i (y_i - f(x_i))^2$$

$$= C$$

— we define $f(x_i)$ as

$$f(x) = f_M(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m)$$

$$- C = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - f_M(x_i))^2$$

$M = 1$ $\quad f_1(x) = f_0(x) + \beta_1 b(x; \gamma_1)$

$$f_m(x) = f_{m-1}(x) + \beta_m b(x; \gamma_m)$$

Example

$f_0(x) = 0 \qquad b(x;\gamma_1) = 1+\gamma x$

optimize

$(\hat{\beta}_m, \hat{\gamma}_m) = \underset{\gamma, \beta}{\arg\min}$

$$\frac{1}{n} \sum_i \left(y_i - f_{m-1}(x_i) - \beta\, b(x_i;\gamma) \right)^2$$

our specific model

$$\frac{1}{n} \sum_i \left(y_i - f_{m-1}(x_i) - \beta(1 + x_i\gamma) \right)^2$$

$$\frac{\partial C}{\partial \beta} = 0 = -\frac{2}{n} \sum_i (1+\gamma x_i)(y_i - \beta(1+\gamma x_i))$$

$$\frac{\partial C}{\partial \gamma} = 0$$

with the first iteration

$f_1(x) = f_0(x) + \beta_1(1+\gamma_1 x)$

$$f_m(x) = f_{m-1}(x) + \beta_m (1+\gamma_m x)$$

continue til $m = M$

This kind of additive expansion is at the heart of many learning techniques ;

— NN ; single hidden layer

$$b(x, y) = \sigma(y_0 + \gamma_1 x)$$

$$\sigma(t) = \frac{1}{1+e^{-t}}$$

γ is set of parameter