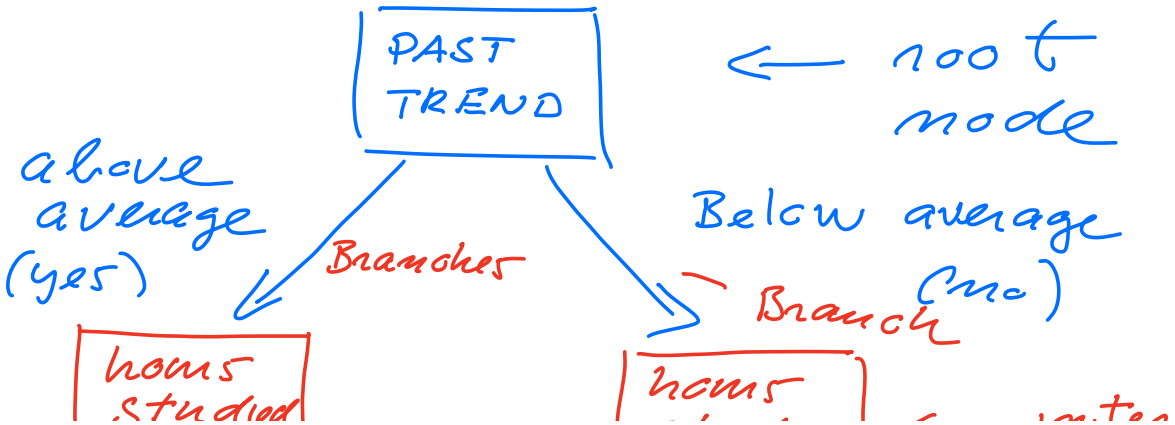
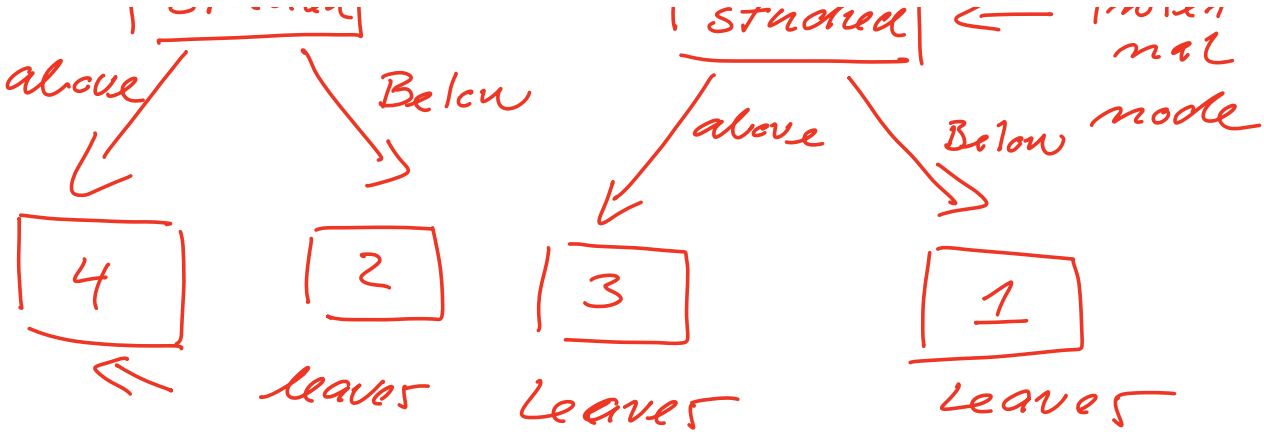
Lecture Jan 14

Classification Tree

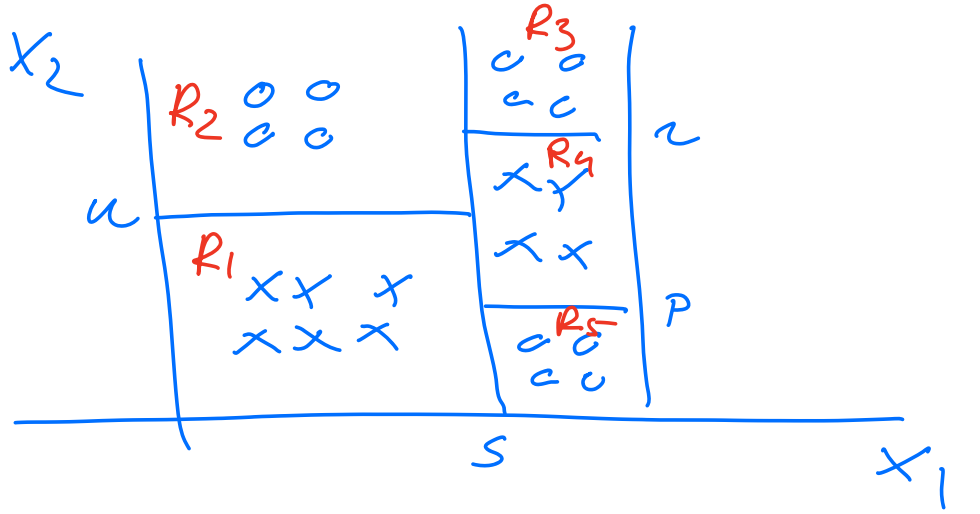
Grade Trend	hours slept	hours studied	G
A (above)	L (ow)	H	A
B (elow)	H (igh)	L	B
A	L	H	A
A	H	H	A
B	L	H	B
A	L	L	B
B	H	H	B
B	L	H	A
A	L	L	B
A	H	H	A

Decision Tree





Regression Tree



if $X_1 \leq S$ and $X_2 \leq u$ R_1

if $X_1 \leq S$ and $X_2 > u$ R_2

if $X_1 > S$ and $X_2 \leq p$ R_5

— , — and $X_2 \leq r$ R_4
and $X_2 > p$

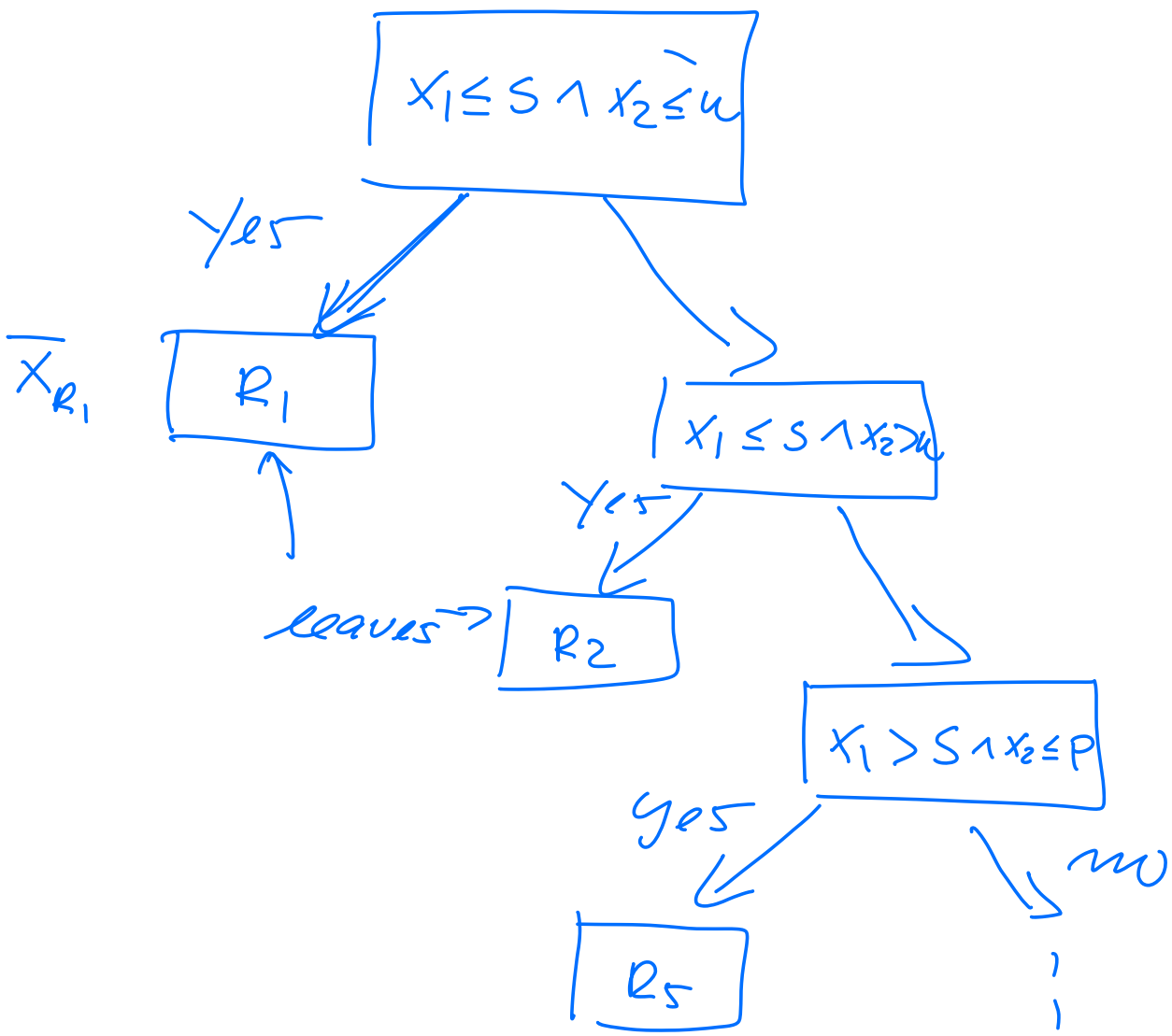
else R_5

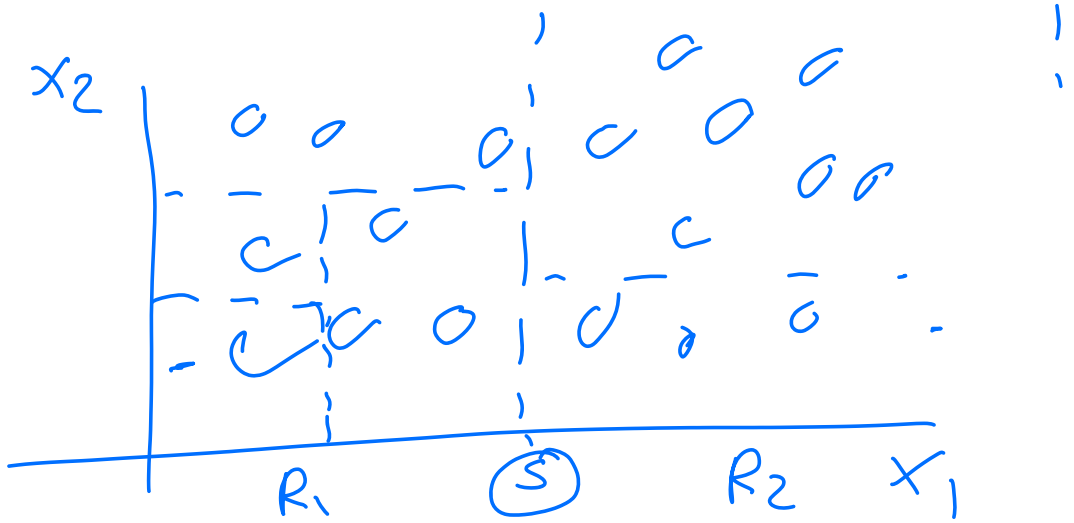
prediction R_1

$x_1 \leq s$ and $x_2 \leq u$

→ average of all observations in R_1

Prediction is average of observations in a domain





MSE(R_1)

MSE(R_2)

Gini calculation

Gini index trend.

$P(\text{above}) = \frac{6}{10} = 3/5$

$P(\text{below}) = 4/10 = 2/5$

if (Trend = A & 6 = A) = $4/6 = 2/3$

if (Trend = A & 6 = B) = $2/6 = 1/3$

$1 - (2/3)^2 - (1/3)^2 = 0.45$

if (Trend = B & 6 = A) = 0

if (Trend = B & G = B) 4/4 = 1

Gini index : 1 - 0 - 1 = 0

Gini index for Past trend

$6/10 \cdot 0.45 + 4/10 \cdot 0 = 0.27$

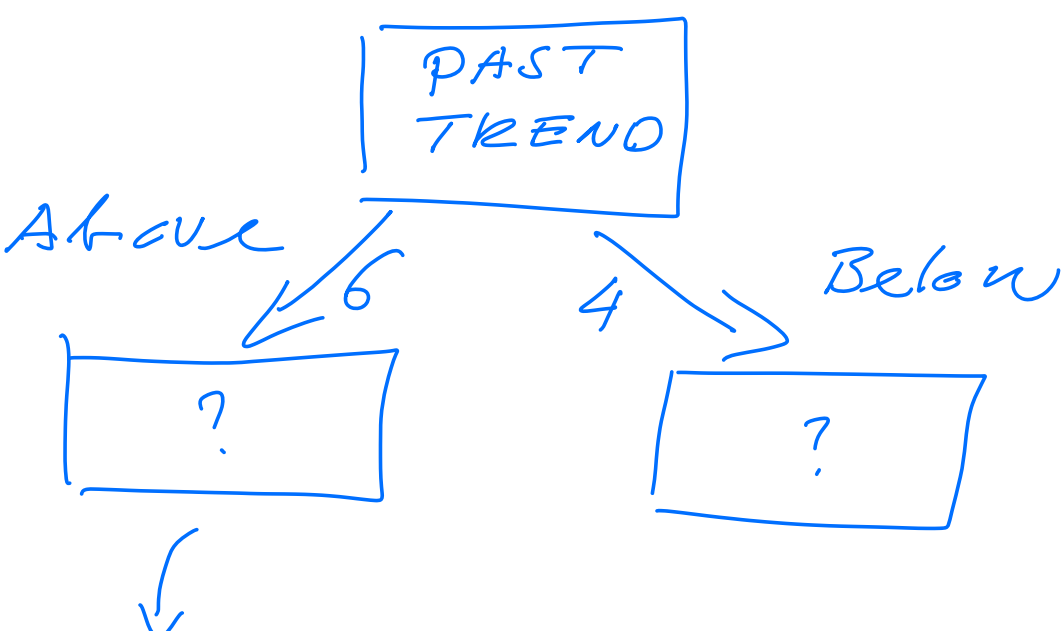
Gini index for hours slept

$g = 0.47$

Gini index for hours studied

$g = 0,34$

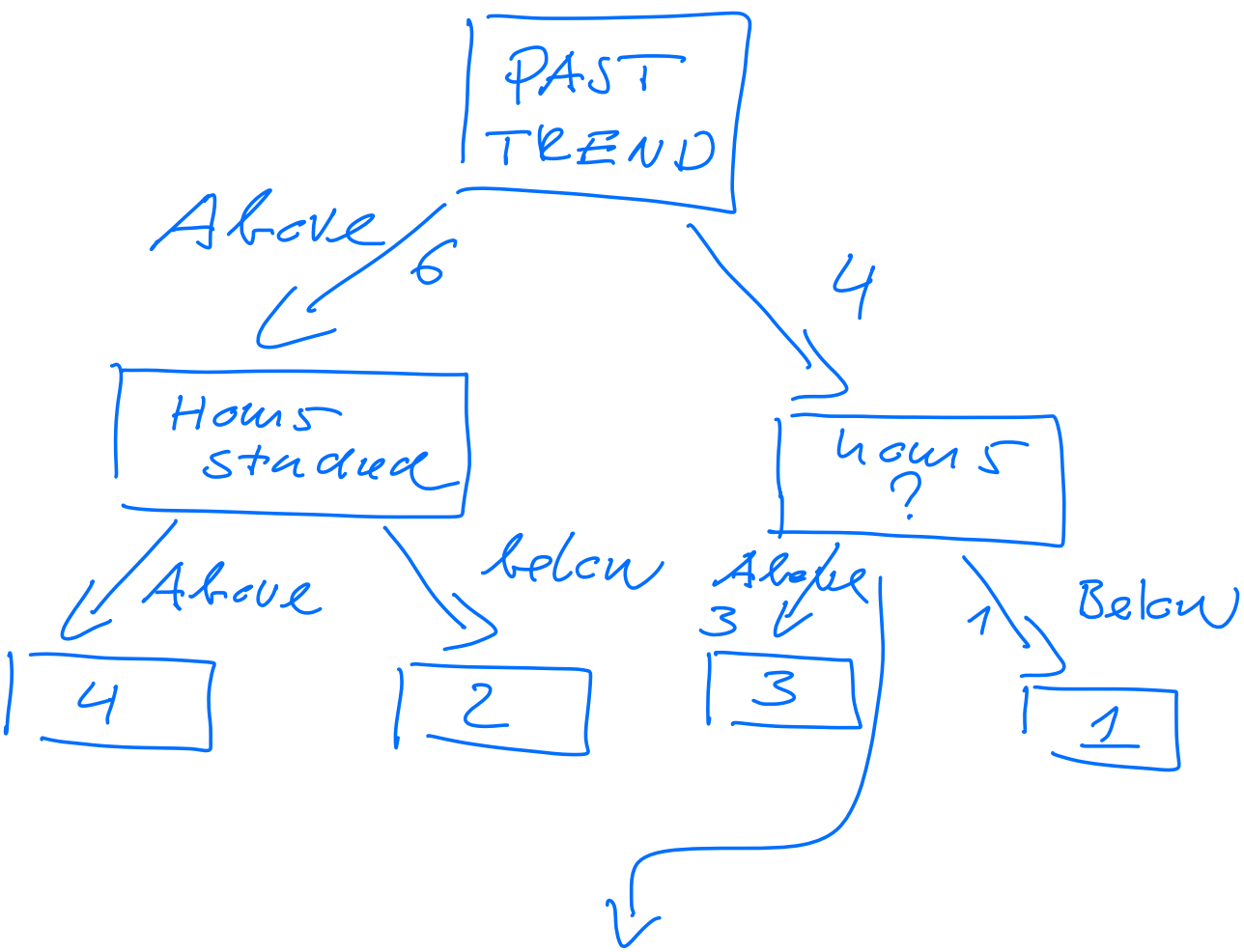
node = past trend



Above	Low	high	above
Above	L	H	A
A	H	H	A
A	L	L	B
A	L	L	B
A	H	H	B

9 - hours studied = 0

9 - - 1 - slept = 0.33



B H L B

B L H B

B H H B

B L H B