

Prosjekttime 3 – HON2200

XGBoost

eXtreme gradient boosted decision trees

Hva er spesielt med XGBoost?

- Bygge spesielle trær vha similarity scores og regularisering.
- “Prune” trær med gain-terskel

Prosedyre for regresjon med XGBoost

- Similarity score & gain. Splitte for å maksimere gain.
- Prune: $\text{gain} < \text{gamma}$?
- Beregne output og nye residualer
- Optional: regularization with Lambda

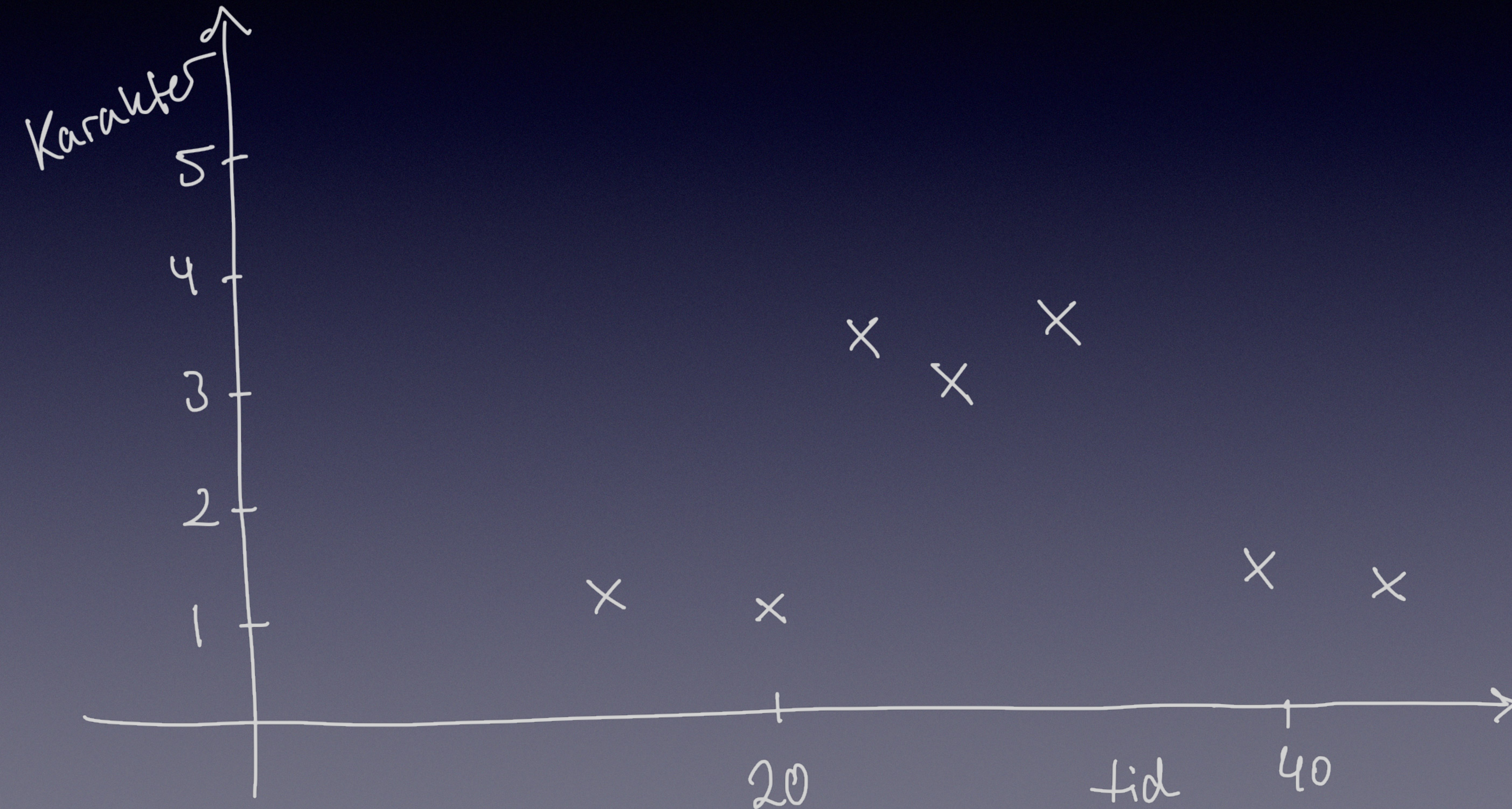
Prosedyre for regresjon med XGBoost

Eksempel

Husket jeg å vise regularisering?

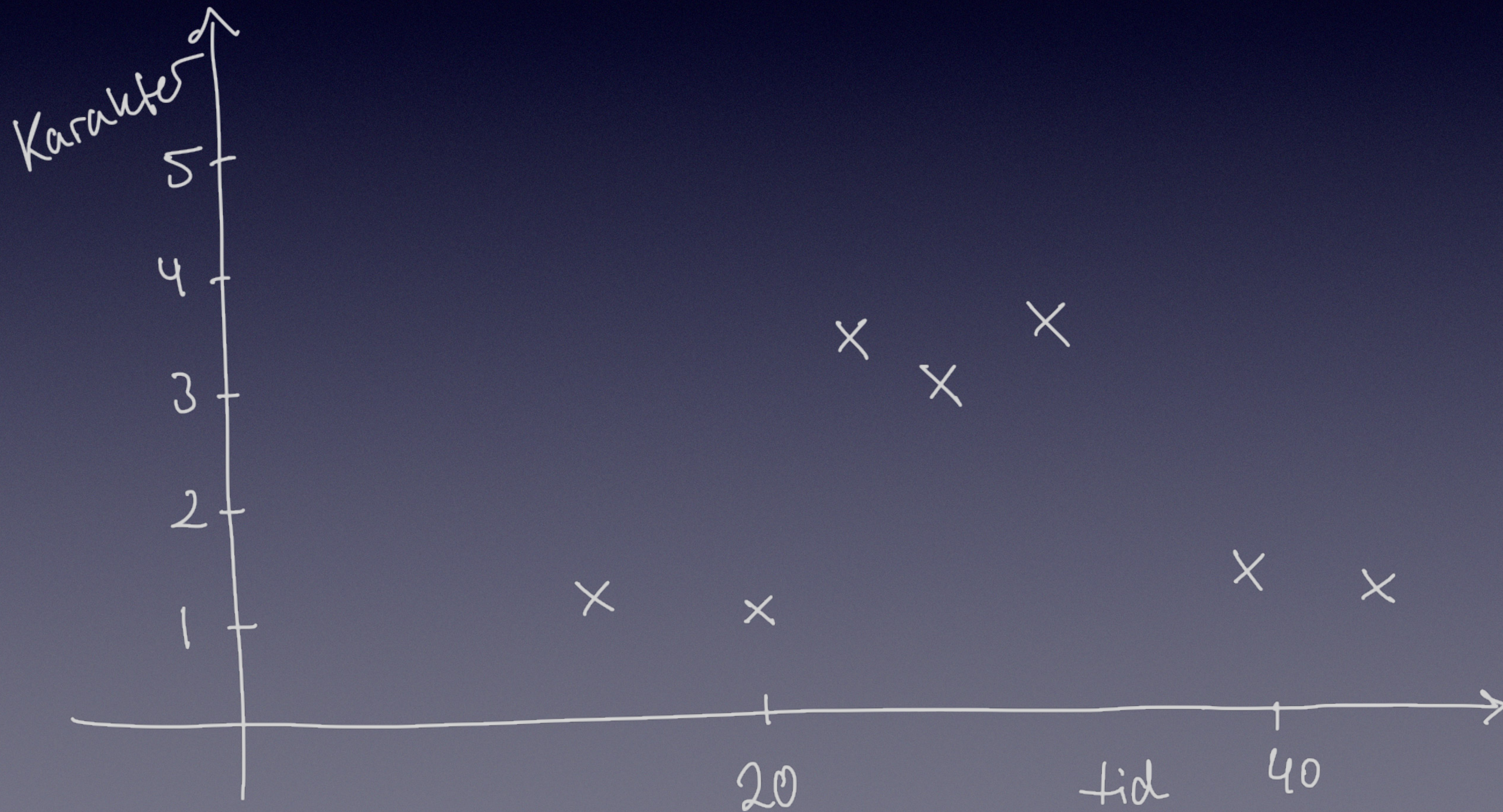
Underveisoppgave

- Finn ut hvilket timetall rotnoden i første XGBoost-tre fra følgende datasett vil dele datasettet basert på.



tid	Karakter
16	1.1
20	1.2
23	3.1
29	2.8
32	3.2
39	1.3
44	1.1

Similarity: $\frac{0,6 + 0,7 + 2,6 + 2,3 + 2,7 + 0,8 + 0,6}{7} \approx 15,2$



tid	Karakter
16	1.1
20	1.2
23	3.1
29	2.8
32	3.2
39	1.3
44	1.1

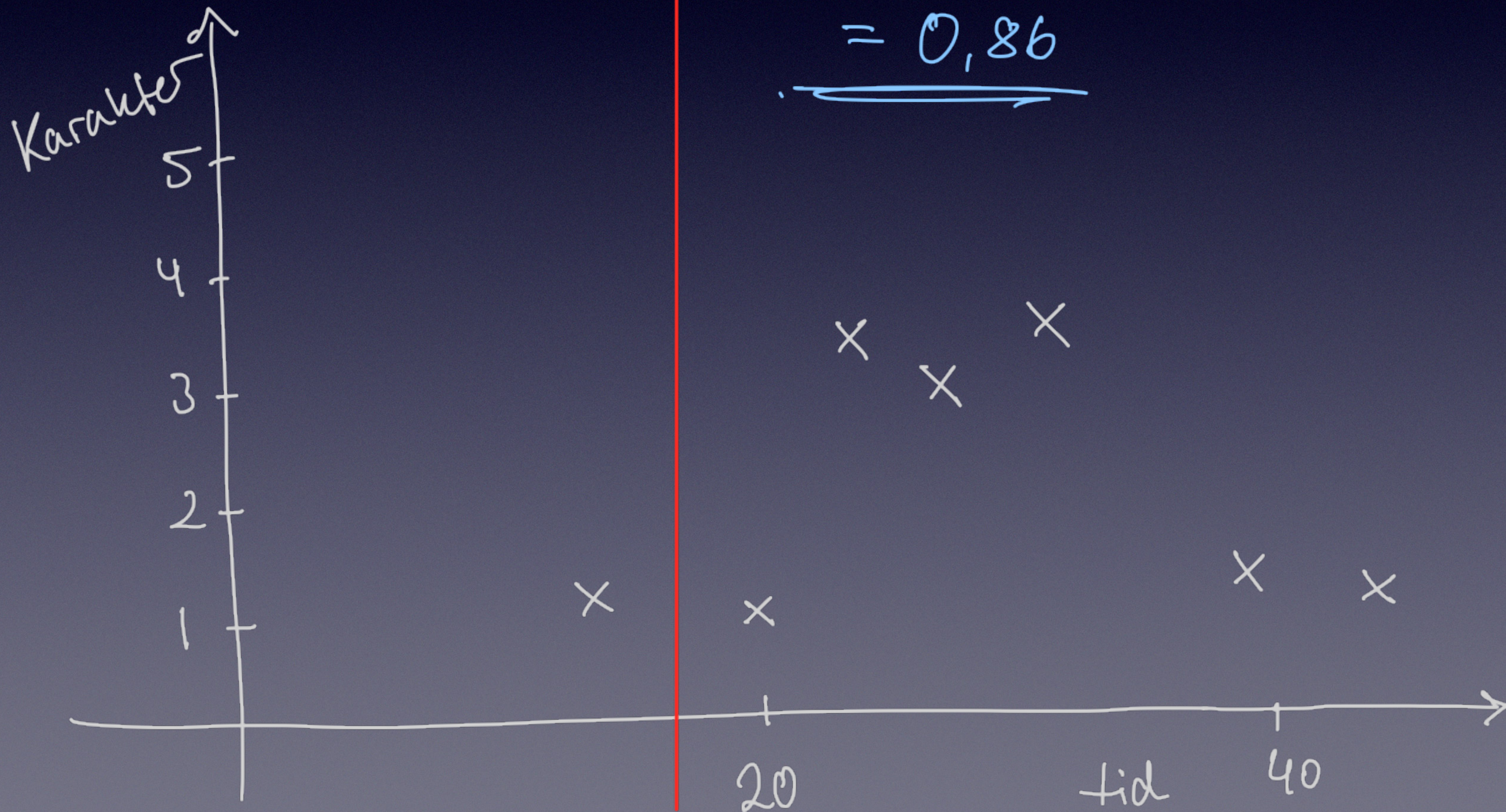
Similarity:

$$\frac{(0,6)^2}{1} = 0,36$$

Similarity!

$$\frac{(0,7 + 2,6 + 2,3 + 2,7 + 0,8 + 0,6)^2}{6} = 15,7$$

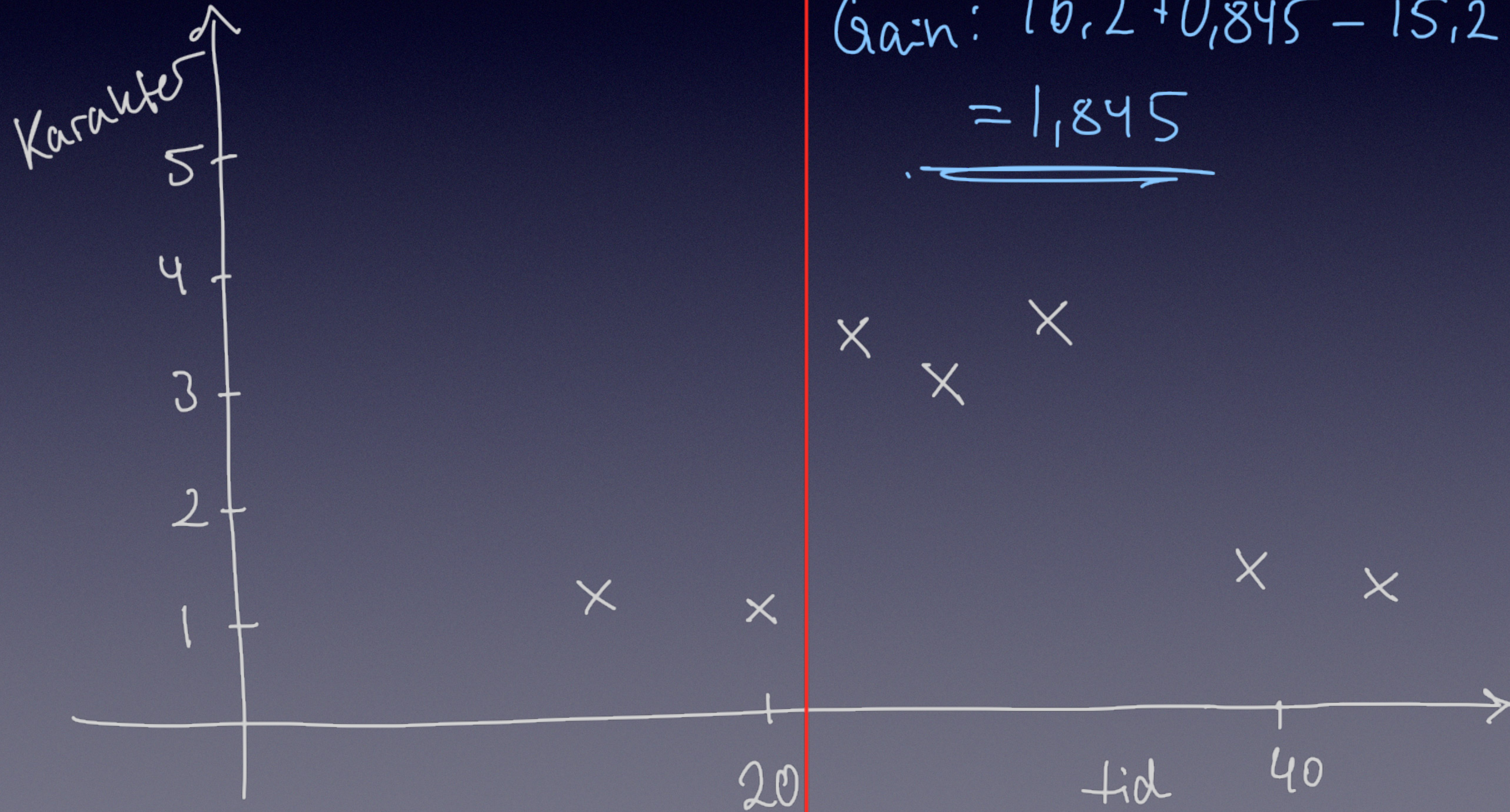
$$\text{Gain: } 15,7 + 0,36 - 15,2 = \underline{\underline{0,86}}$$



tid	Karakter
16	1.1
20	1.2
23	3.1
29	2.8
32	3.2
39	1.3
44	1.1

Similarity:

$$\frac{(0,6+0,7)^2}{2} = 0,845$$

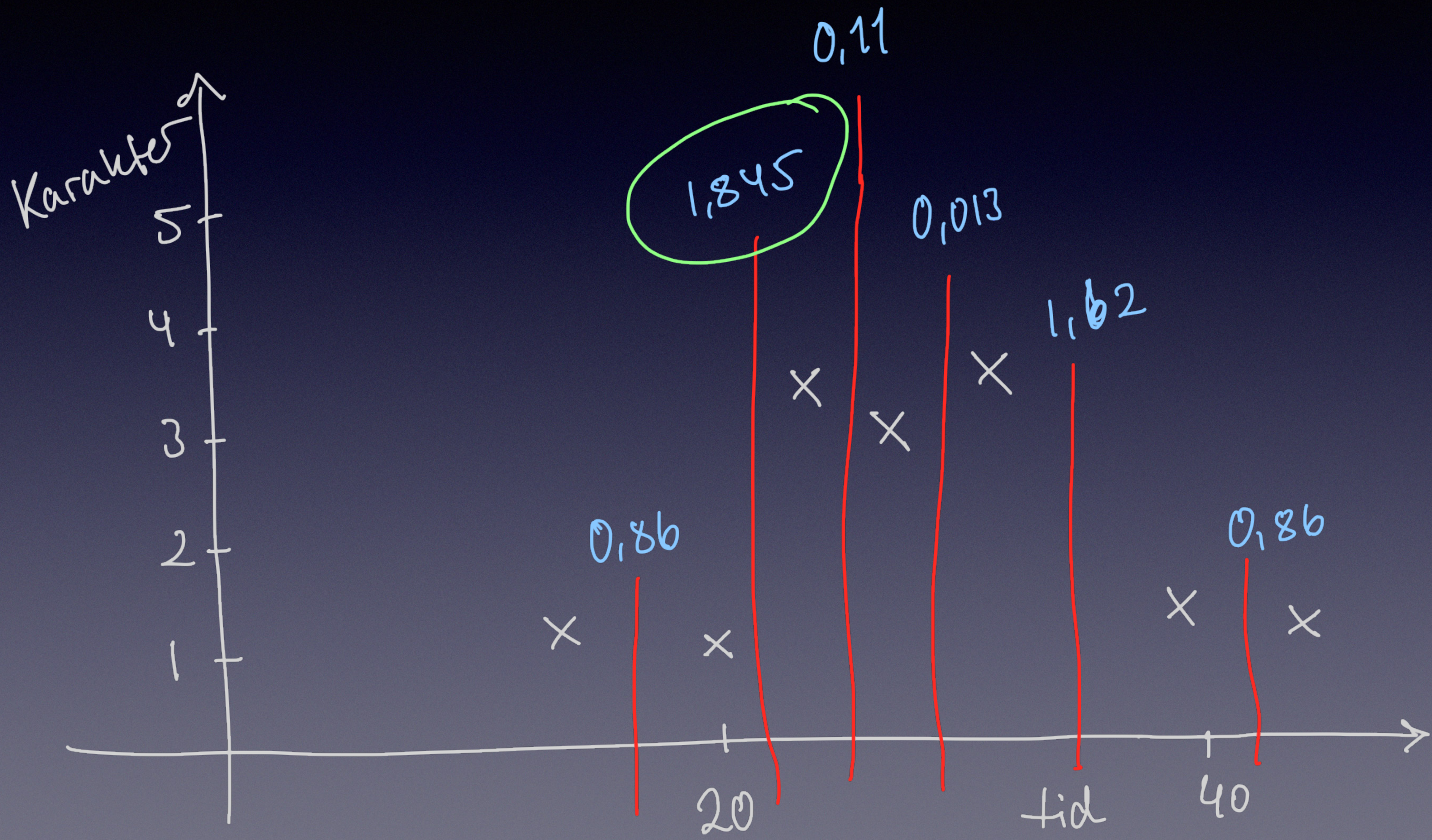


Similarity!

$$\frac{(2,6+2,3+2,7+0,8+0,6)^2}{6} = 16,2$$

$$\text{Gain: } 16,2 + 0,845 - 15,2 = 1,845$$

tid	Karakter
16	1.1
20	1.2
23	3.1
29	2.8
32	3.2
39	1.3
44	1.1



tid	Karakter
16	1.1
20	1.2
23	3.1
29	2.8
32	3.2
39	1.3
44	1.1

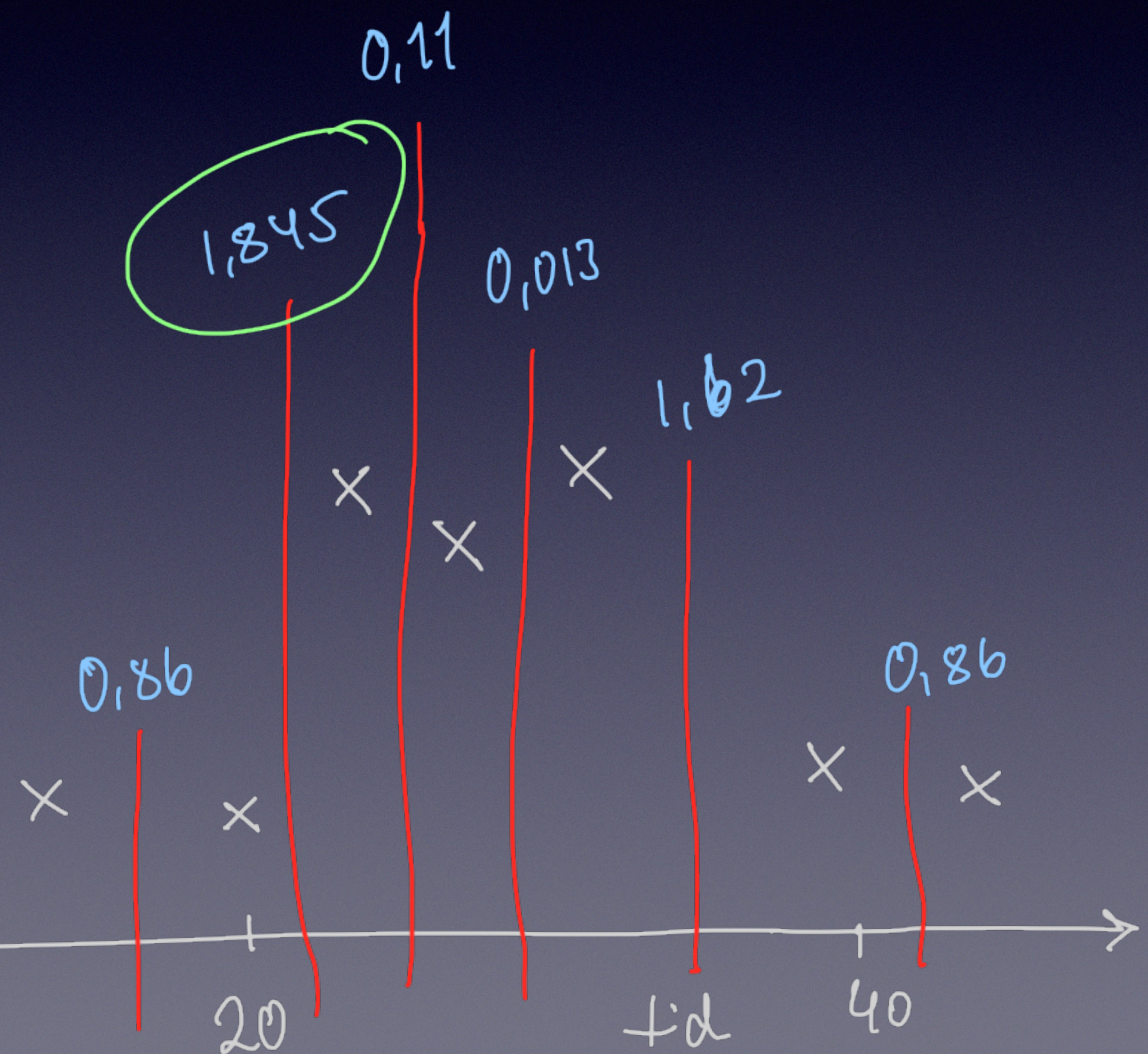
Tid < 21,5

0,65

...

Karakter

5
4
3
2
1



tid	Karakter
16	1.1
20	1.2
23	3.1
29	2.8
32	3.2
39	1.3
44	1.1