Collectieve Intelligentie

Deel 2: KNN en Similarities

Recommender systems

Machine learning algorithms

Vaak veel wiskunde

Recommender systems

Machine learning algorithms

Vaak veel wiskunde

Classificatie vs degressie

130^2 m

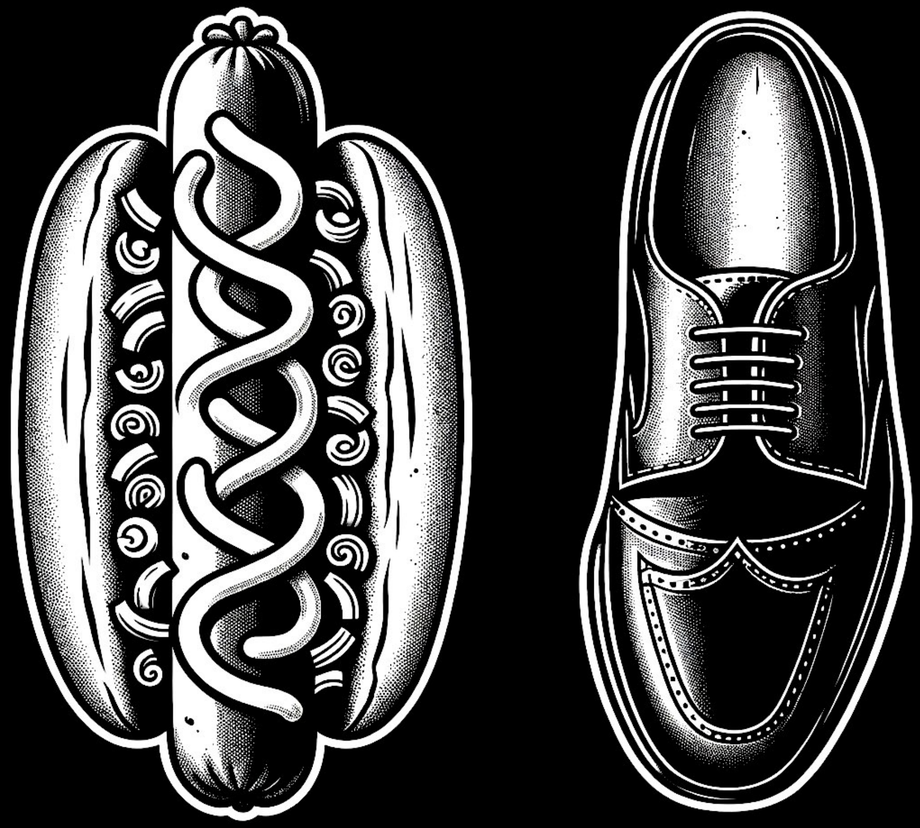
5 kamers

2 badkamers

Centrum

1 295 000

Regressie

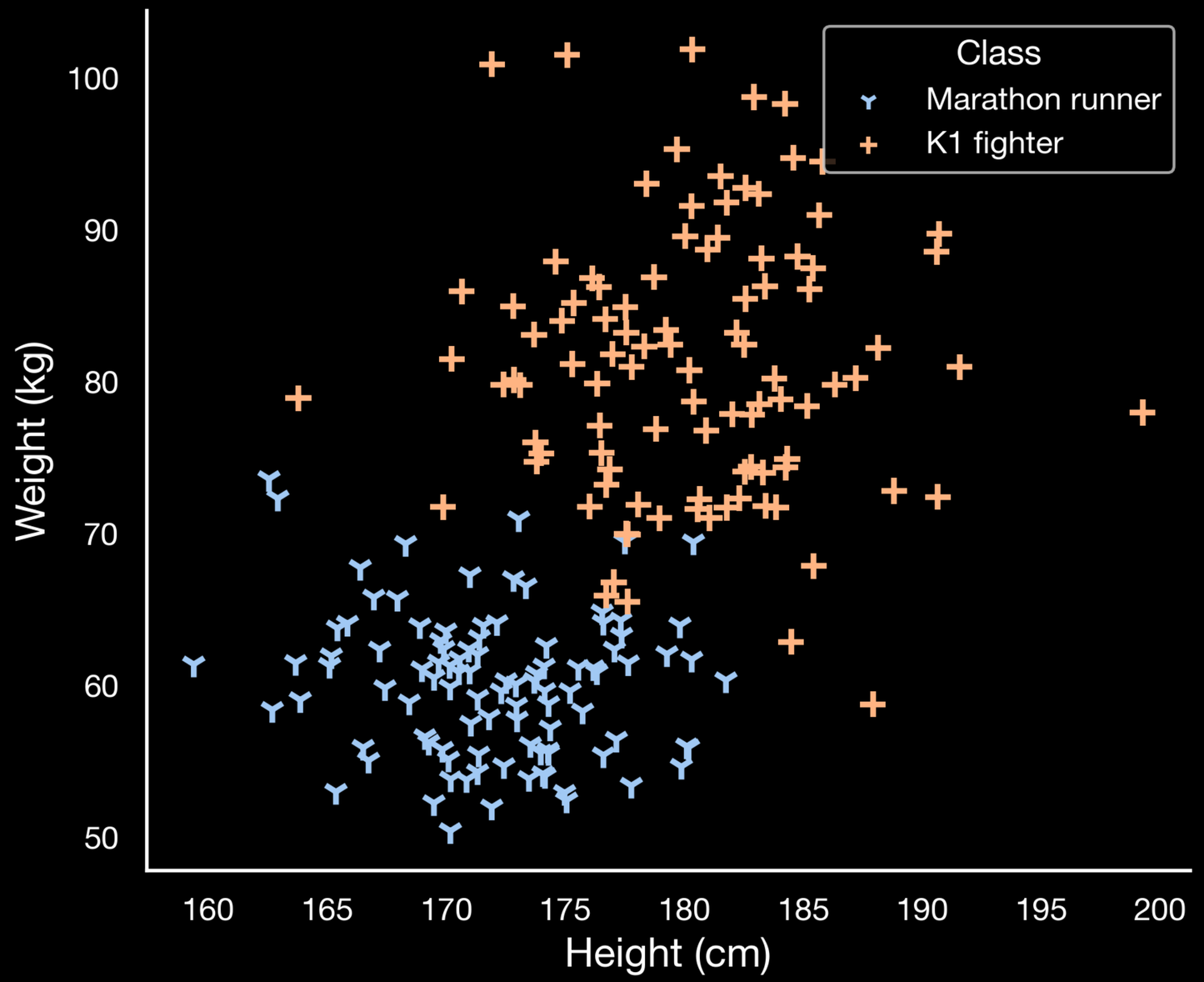


hotdog! not a hotdog!

Classificatie

Harry

Marathonloper of K1-vechter?

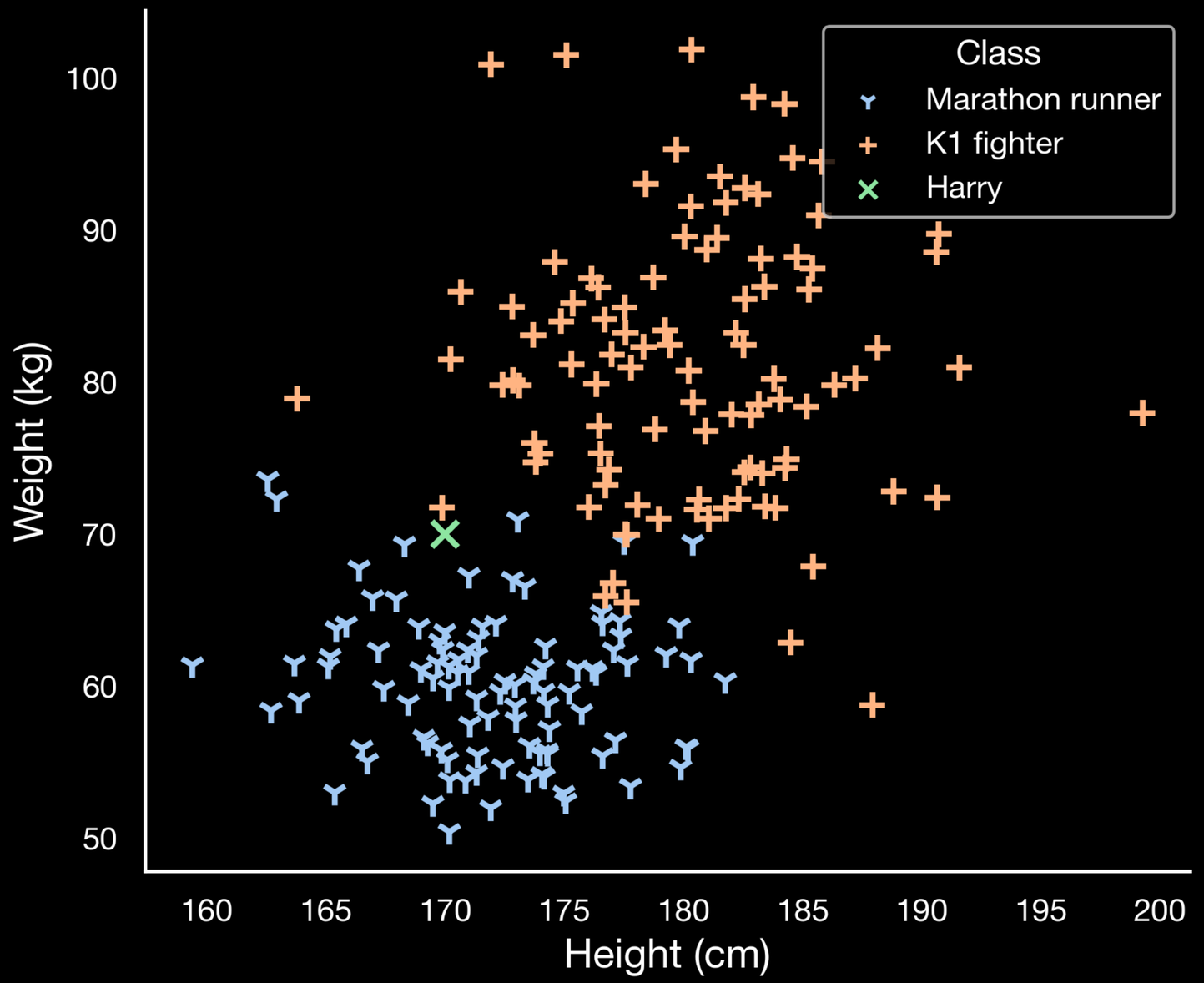


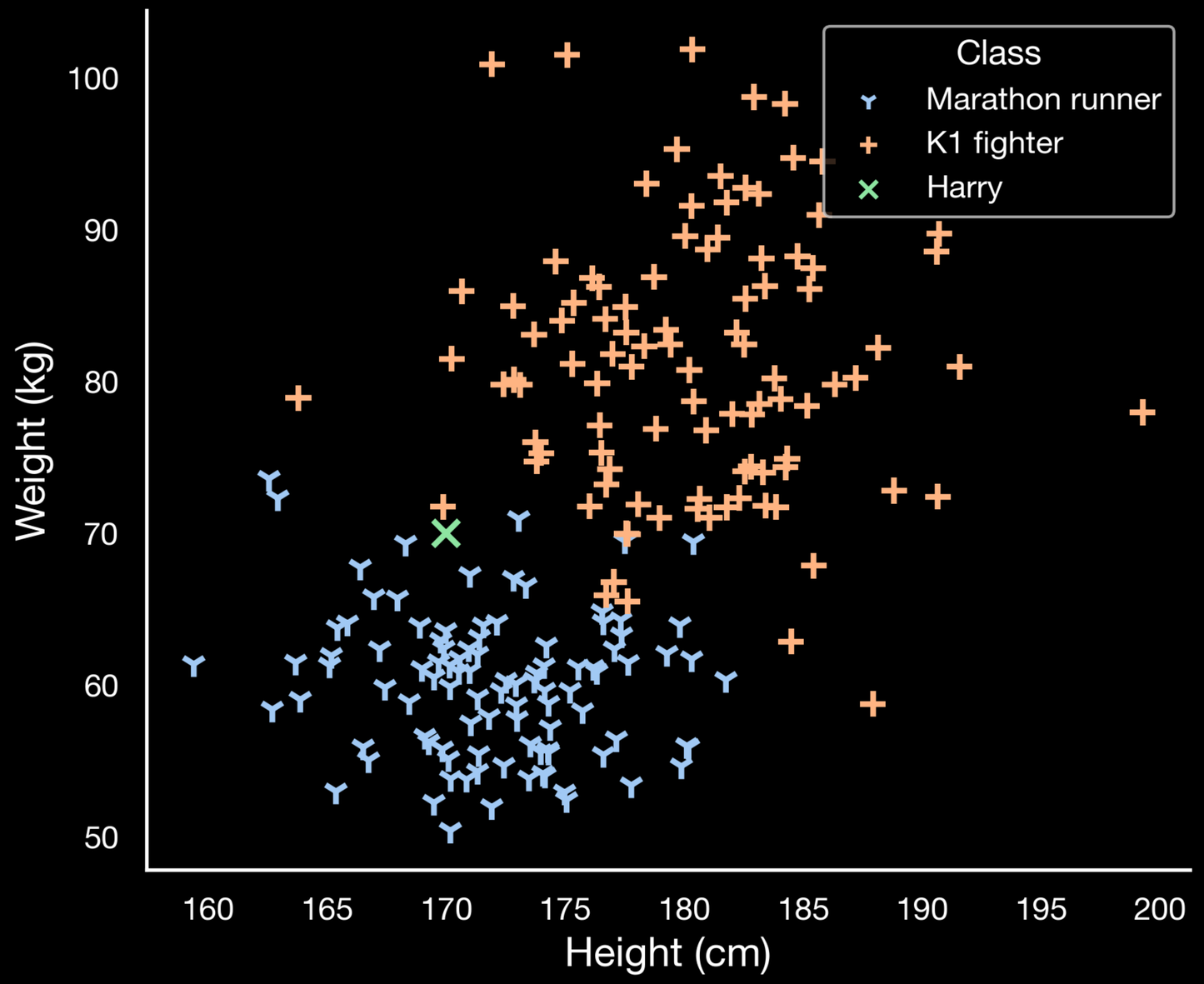
KNN - Marathon vs K1

Standaard *machine learning* classificatie-algoritme

KNN - Marathon vs K1

Standaard *machine learning* classificatie-algoritme

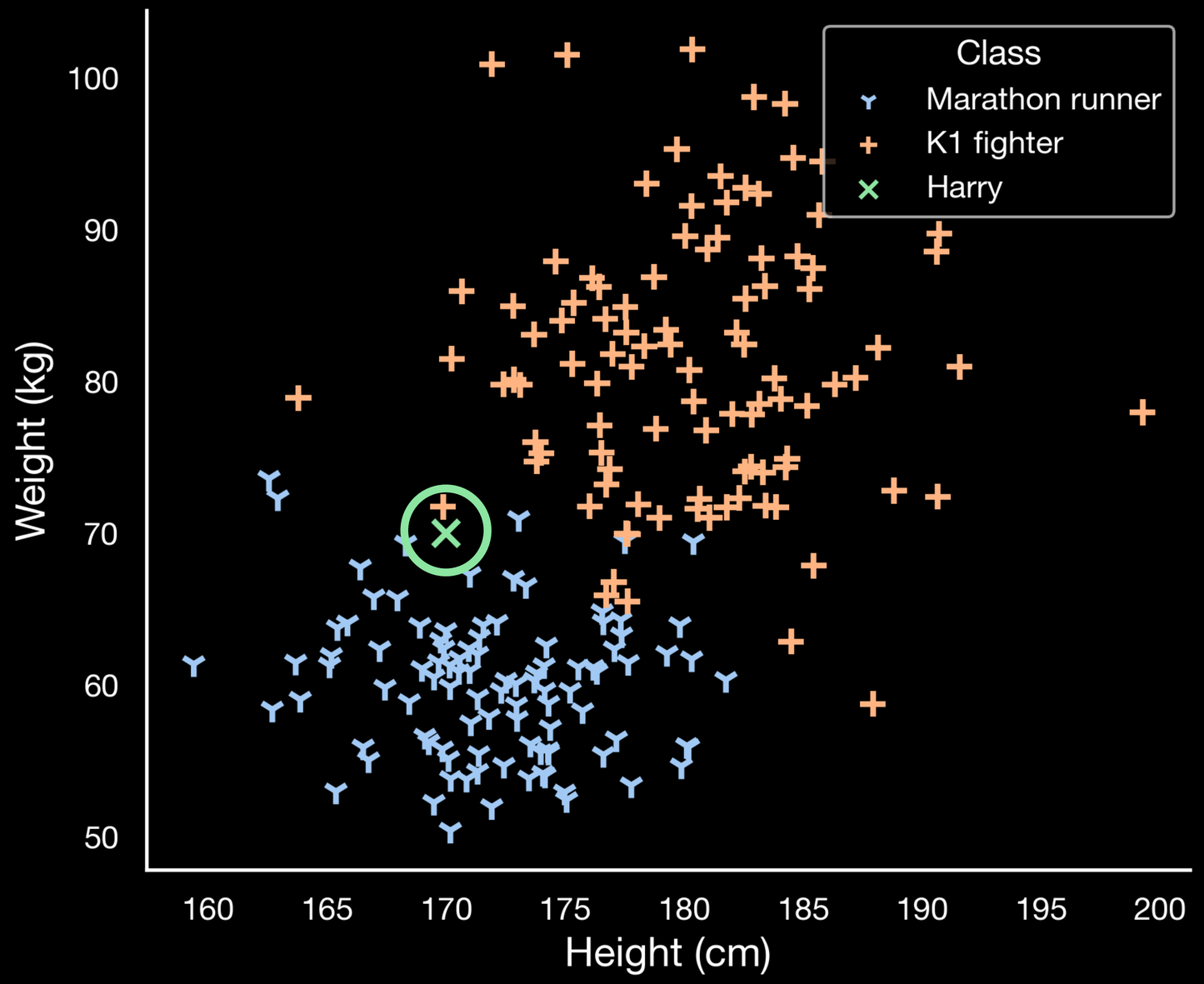




KNN - Marathon vs K1

Standaard *machine learning* classificatie-algoritme

Height (cm)	Weight (kg)	Class	Distance (to Harry)
169.8	71.7	K1	1.7495
168.3	69.3	Marathon	1.8240
171.0	67.2	Marathon	2.9112
173.0	70.9	Marathon	3.1996
172.8	67.0	Marathon	4.1192
...	...	...	...
171.9	100.9	K1	30.9844
182.9	98.7	K1	31.5391
184.2	98.3	K1	31.6973
175.0	101.5	K1	31.9403
180.3	101.8	K1	33.5348



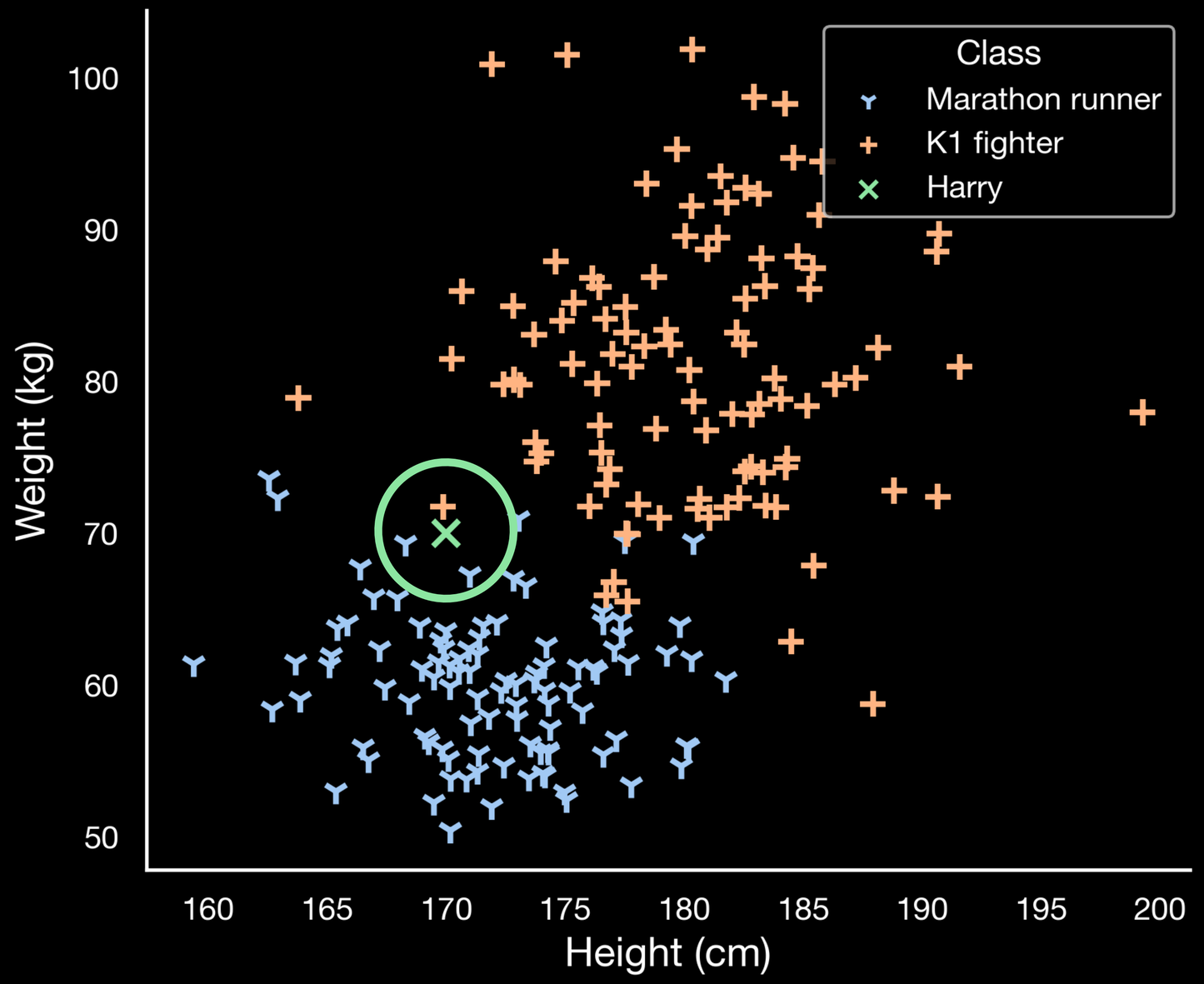
KNN - Marathon vs K1

Standaard *machine learning* classificatie-algoritme

Aanpak: **1** *Nearest Neighbor*

Probleem: ruis/onbetrouwbaar

Height (cm)	Weight (kg)	Class	Distance (to Harry)
169.8	71.7	K1	1.7495
168.3	69.3	Marathon	1.8240
171.0	67.2	Marathon	2.9112
173.0	70.9	Marathon	3.1996
172.8	67.0	Marathon	4.1192
...	...	...	...
171.9	100.9	K1	30.9844
182.9	98.7	K1	31.5391
184.2	98.3	K1	31.6973
175.0	101.5	K1	31.9403
180.3	101.8	K1	33.5348

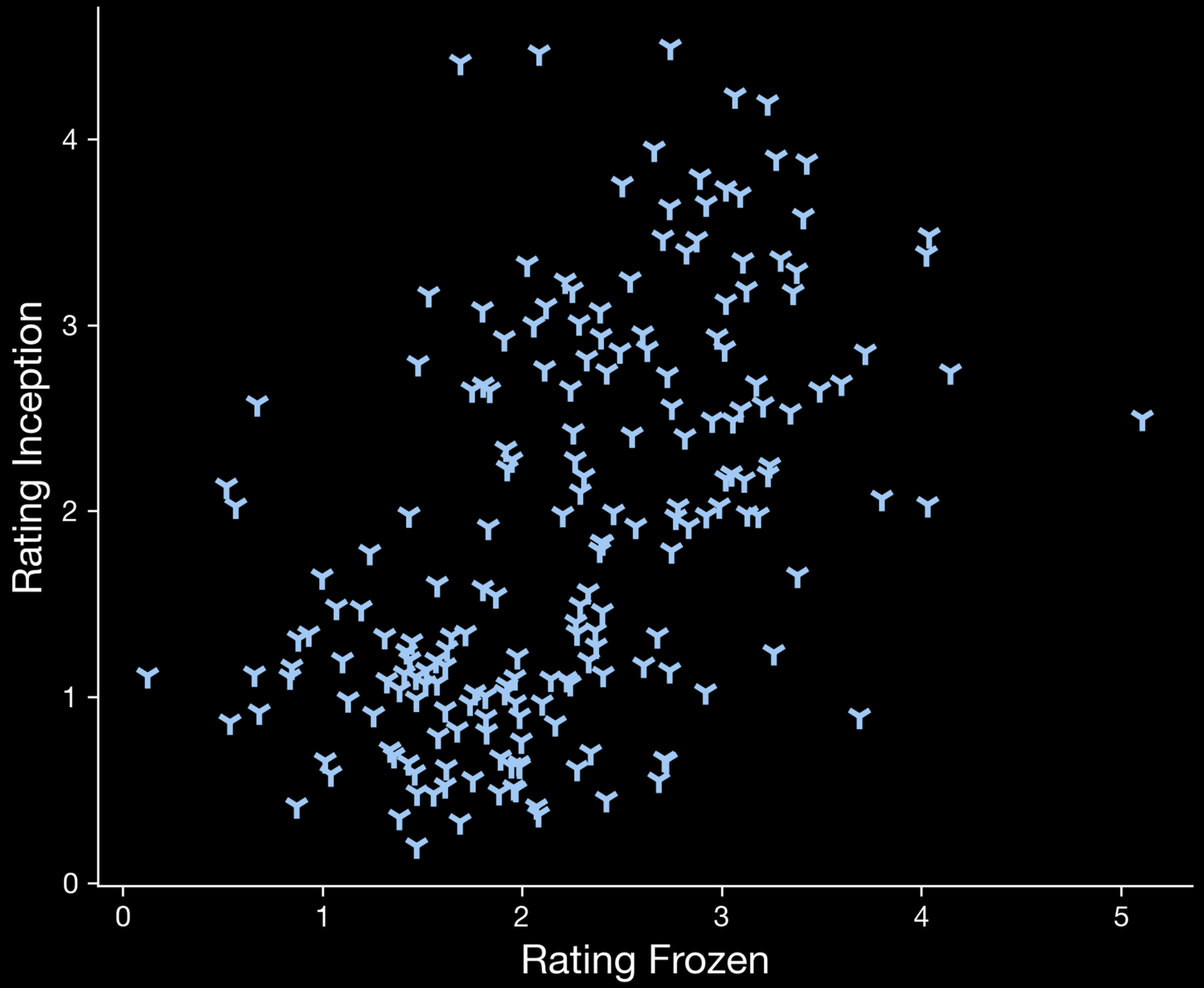


KNN - Marathon vs K1

Standaard *machine learning* classificatie-algoritme

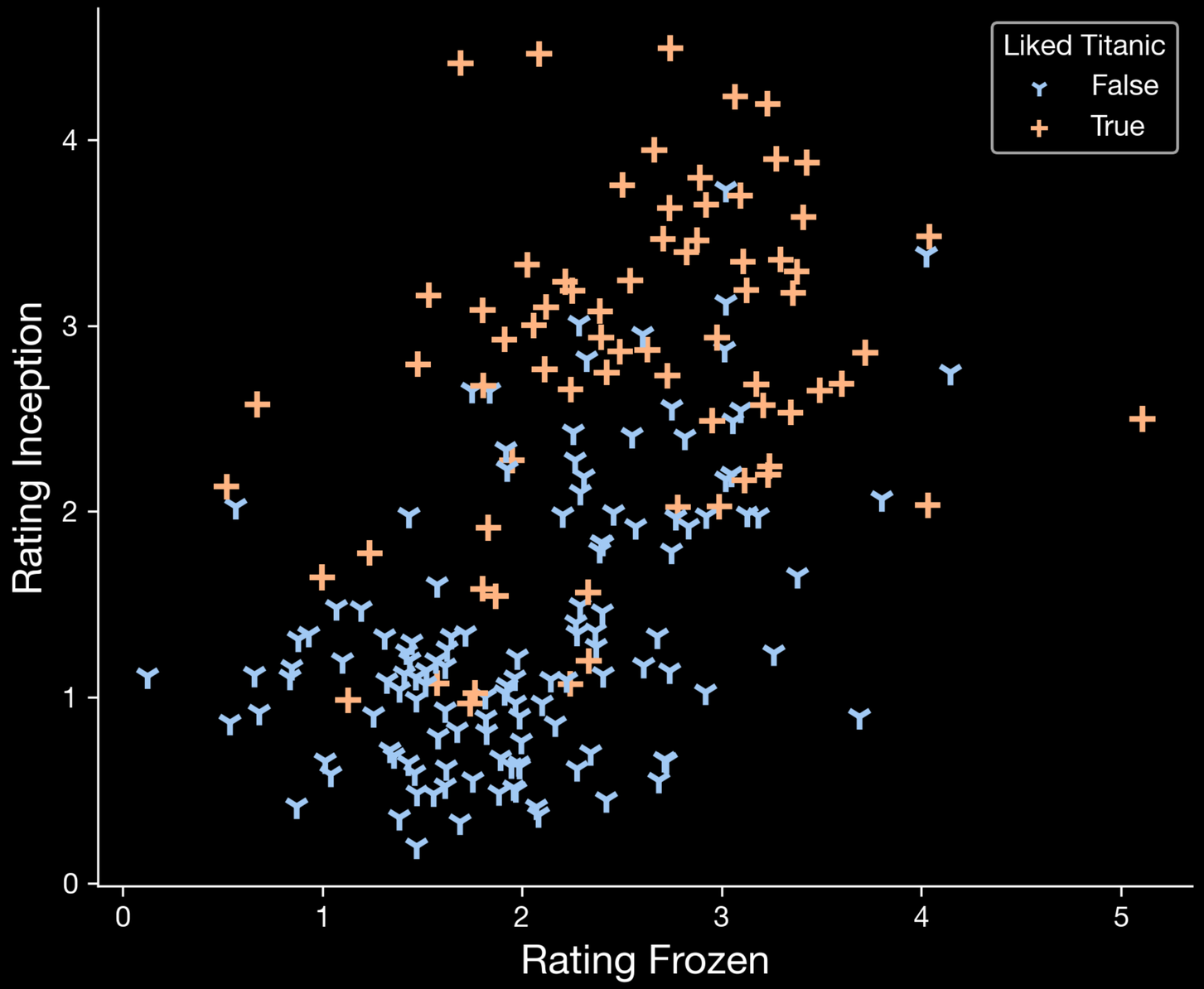
Oplossing: **K** *Nearest Neighbor*

Height (cm)	Weight (kg)	Class	Distance (to Harry)
169.8	71.7	K1	1.7495
168.3	69.3	Marathon	1.8240
171.0	67.2	Marathon	2.9112
173.0	70.9	Marathon	3.1996
172.8	67.0	Marathon	4.1192
...	...	...	...
171.9	100.9	K1	30.9844
182.9	98.7	K1	31.5391
184.2	98.3	K1	31.6973
175.0	101.5	K1	31.9403
180.3	101.8	K1	33.5348



KNN - Movies

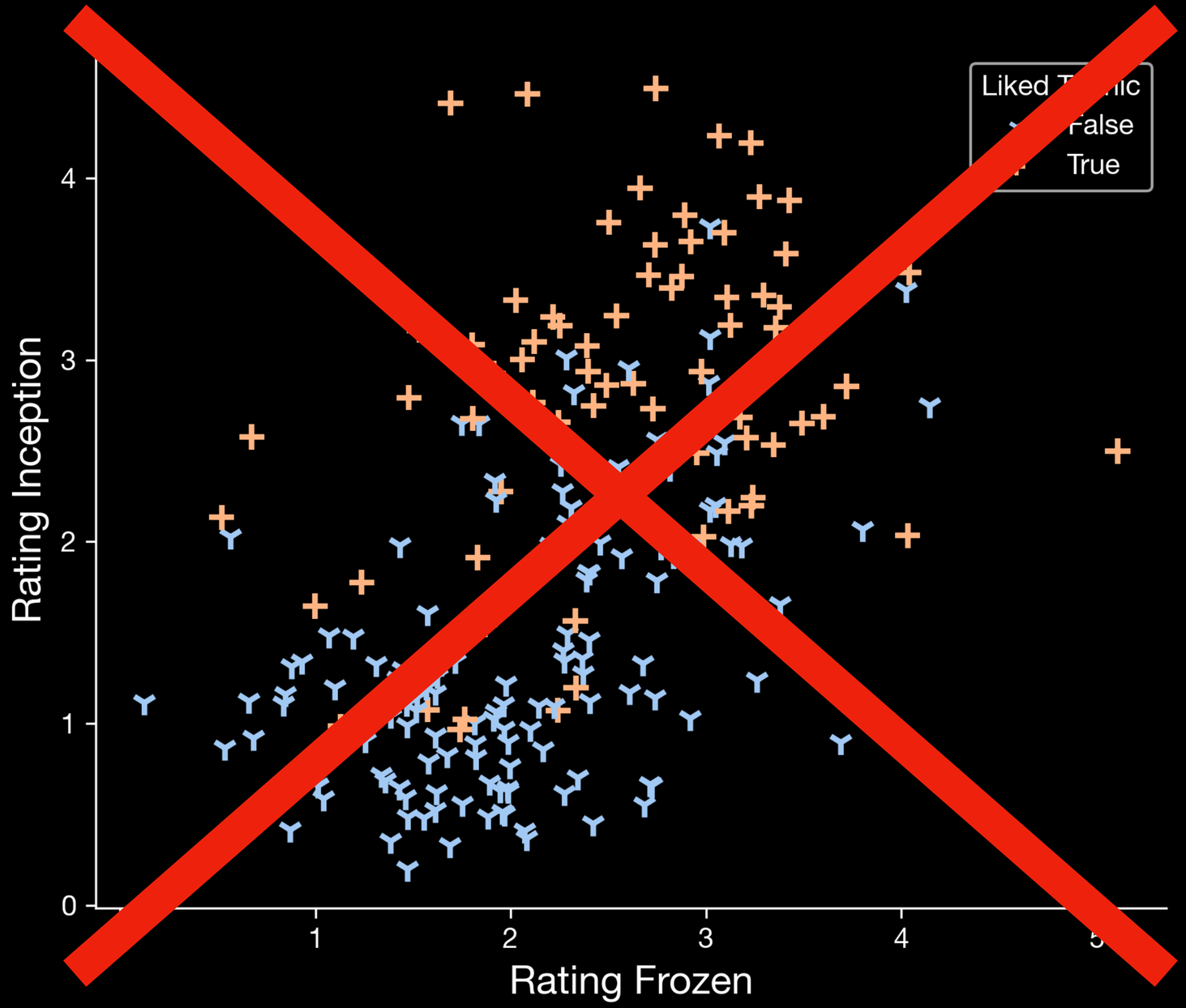
"Standaard" *machine learning* classificatie-algoritme



KNN - Movies

"Standaard" *machine learning* classificatie-algoritme

Wat we zouden willen

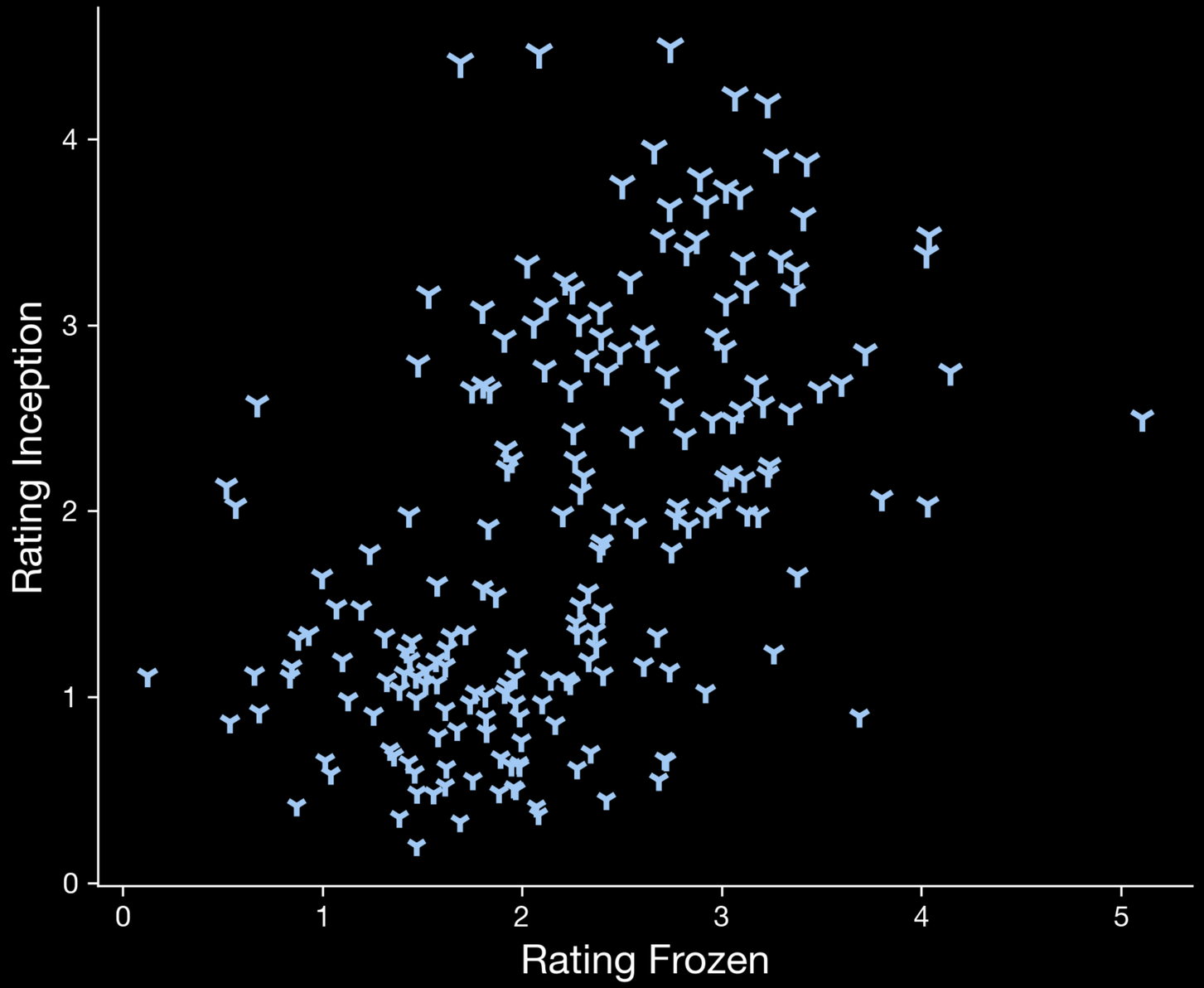


KNN - Movies

"Standaard" *machine learning* classificatie-algoritme

Classificatie niet mogelijk

Data bevat geen *class*-informatie

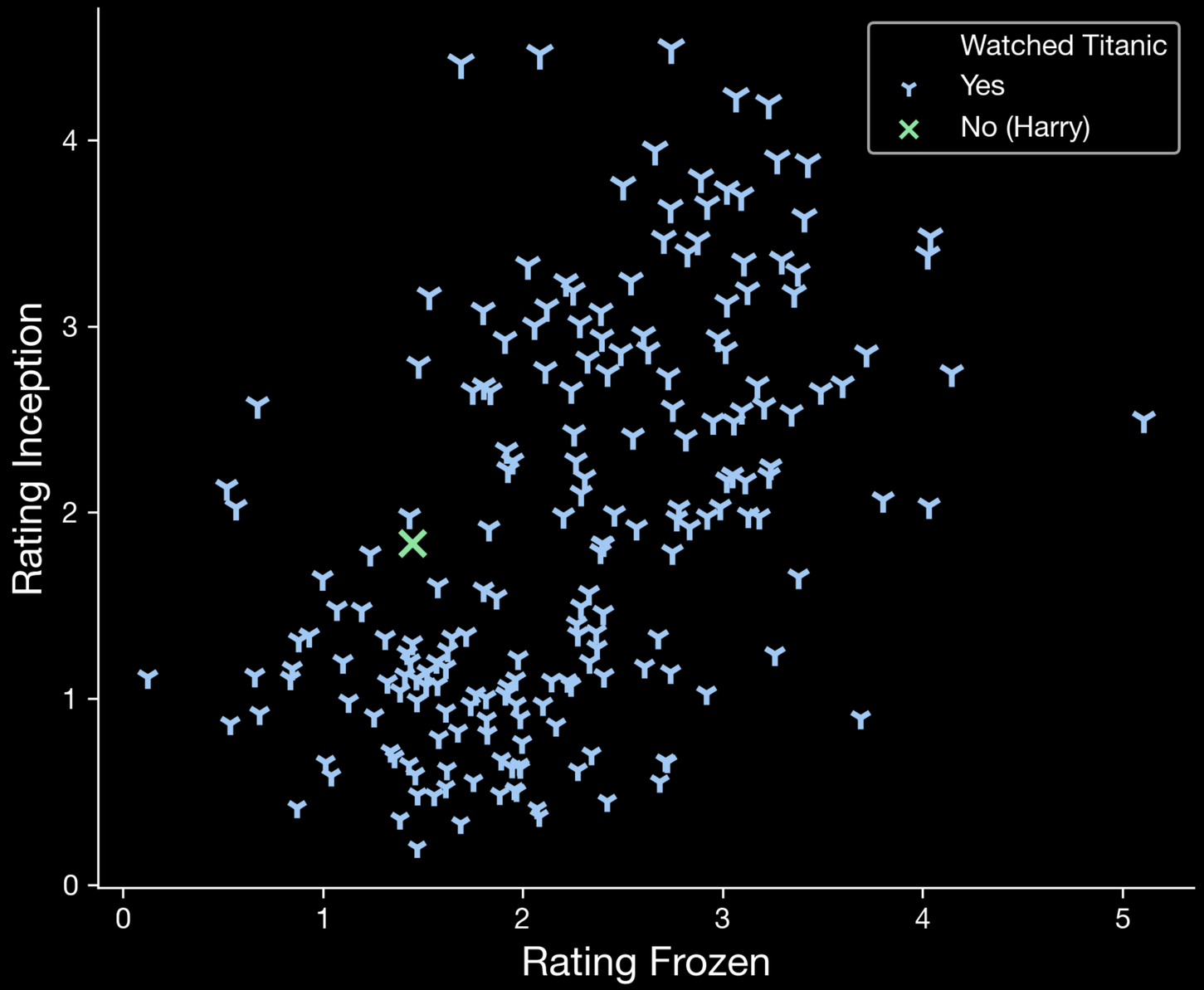


KNN - Movies

"Standaard" *machine learning* ~~classificatie~~ regressie-algoritme

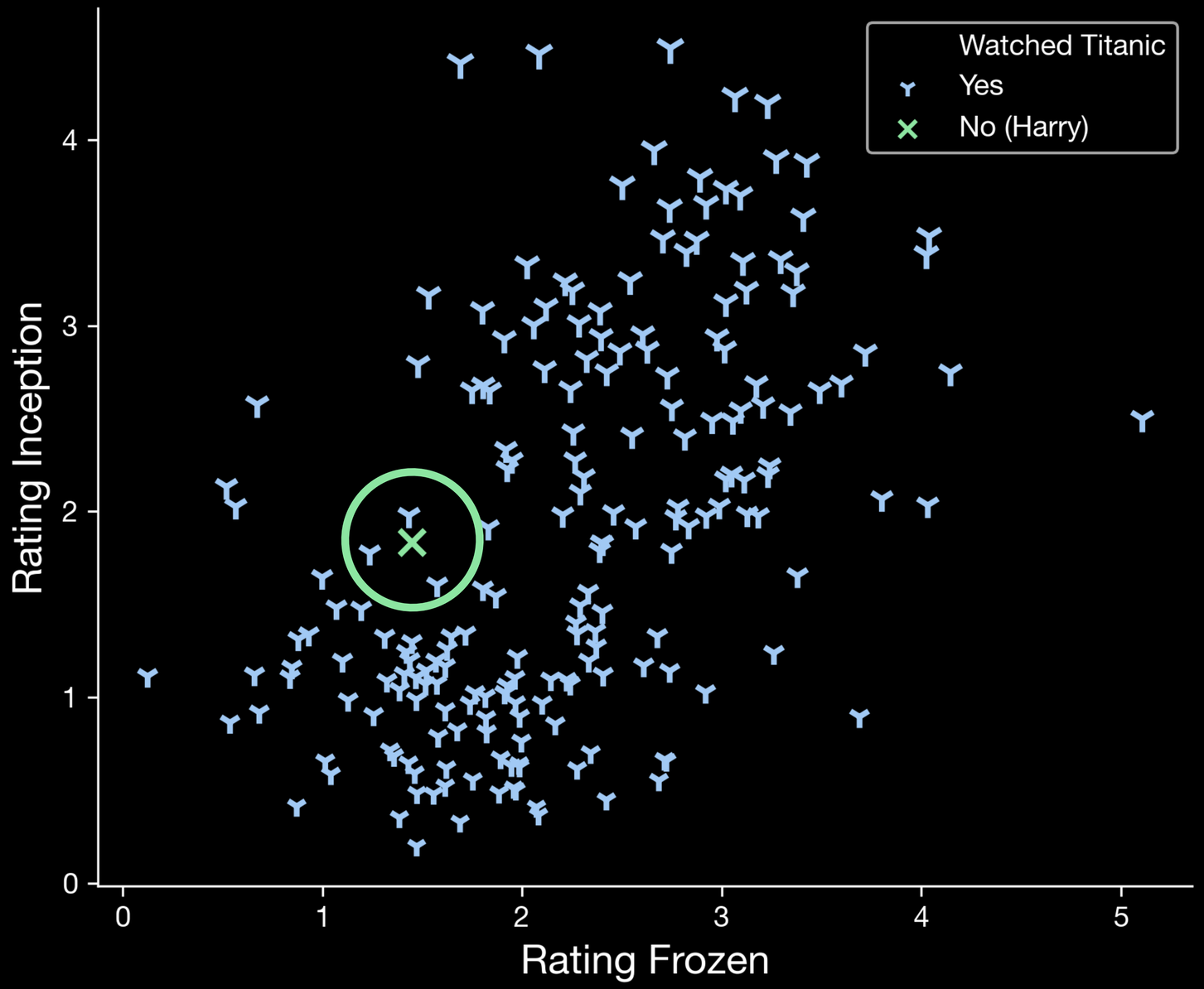
Markers:

- rating Titanic = 5.0
- rating Titanic = 0.0



KNN - Movies

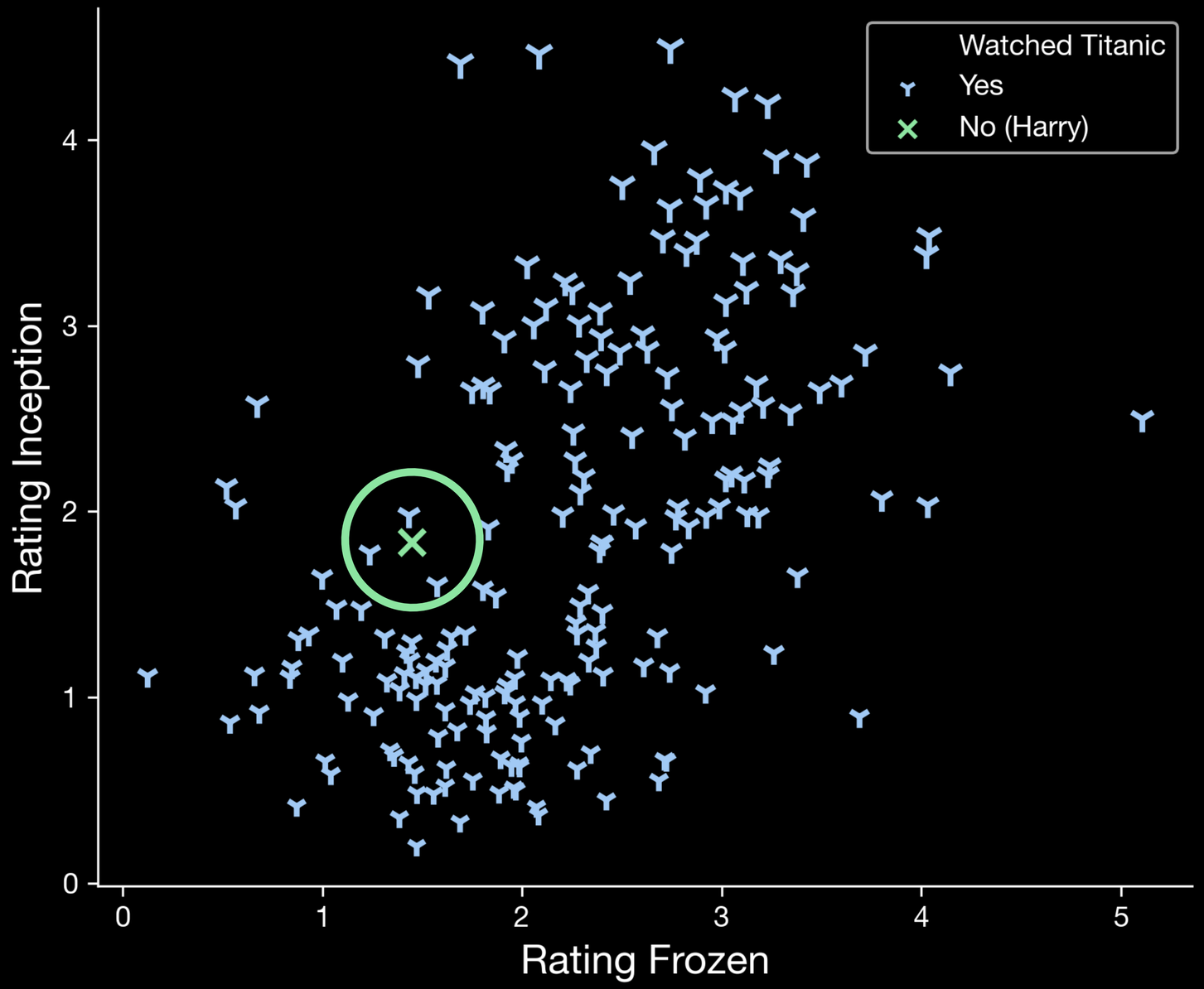
Standaard *machine learning* classificatie-algoritme



KNN - Movies

Standaard *machine learning* classificatie-algoritme

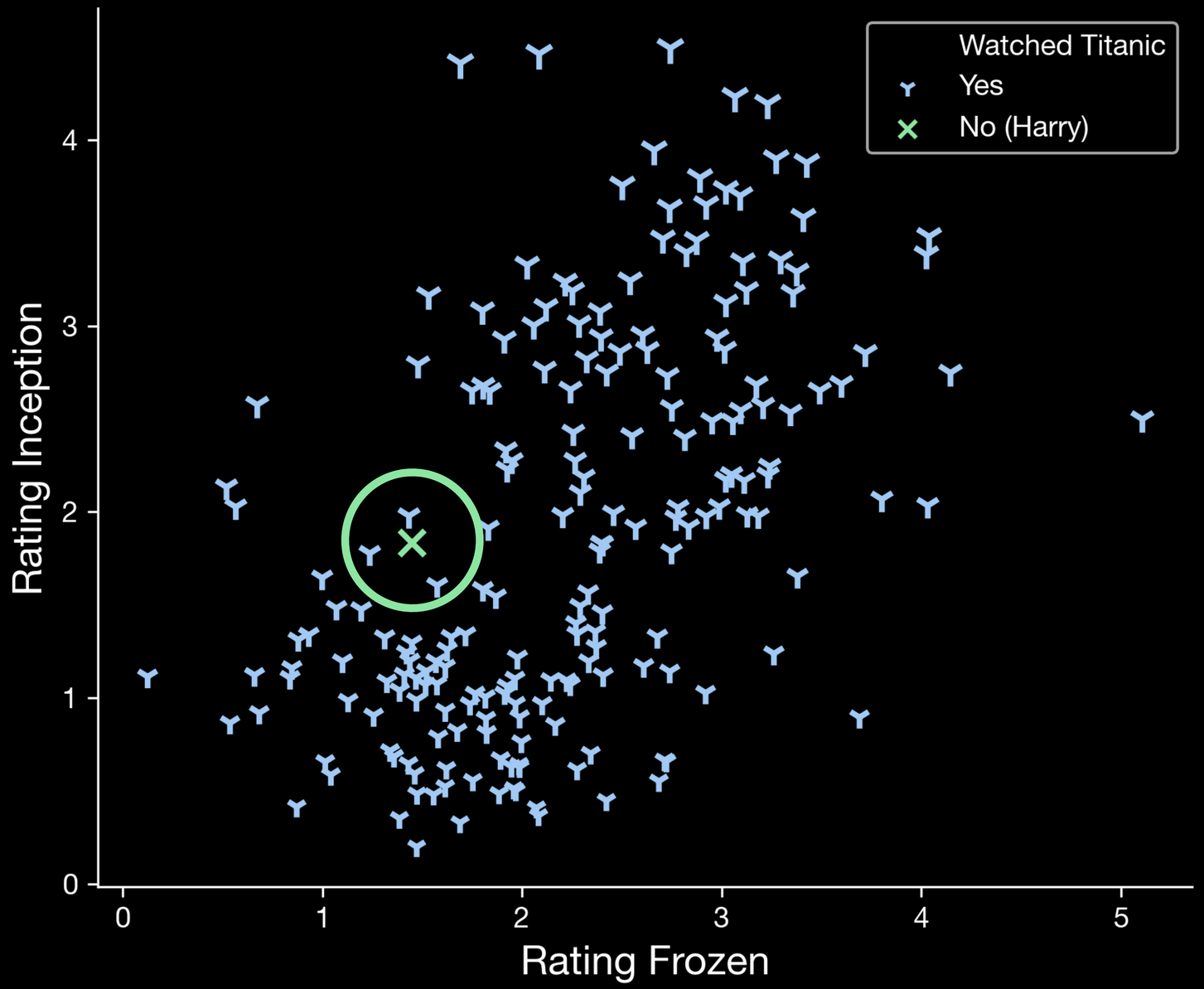
Rating (Froze)	Rating (Incep)	Rating (Titan)	Distance (to Harry)
1.434	1.979	2.014	0.1750
1.238	1.777	2.629	0.1824
1.576	1.606	1.382	0.2911
1.832	1.913	2.797	0.3200
1.805	1.584	3.815	0.4119
...	...	...	...
1.692	4.410	4.252	3.0984
3.068	4.230	2.711	3.1539
3.231	4.193	3.161	3.1697
2.087	4.461	4.793	3.1940
2.744	4.492	2.959	3.3535



KNN - Movies

Standaard *machine learning* classificatie-algoritme

Rating (Froze)	Rating (Incep)	Rating (Titan)	Distance (to Harry)
1.434	1.979	2.014	0.1750
1.238	1.777	2.629	0.1824
1.576	1.606	1.382	0.2911
1.832	1.913	2.797	0.3200
1.805	1.584	3.815	0.4119
...	...	...	...
1.692	4.410	4.252	3.0984
3.068	4.230	2.711	3.1539
3.231	4.193	3.161	3.1697
2.087	4.461	4.793	3.1940
2.744	4.492	2.959	3.3535



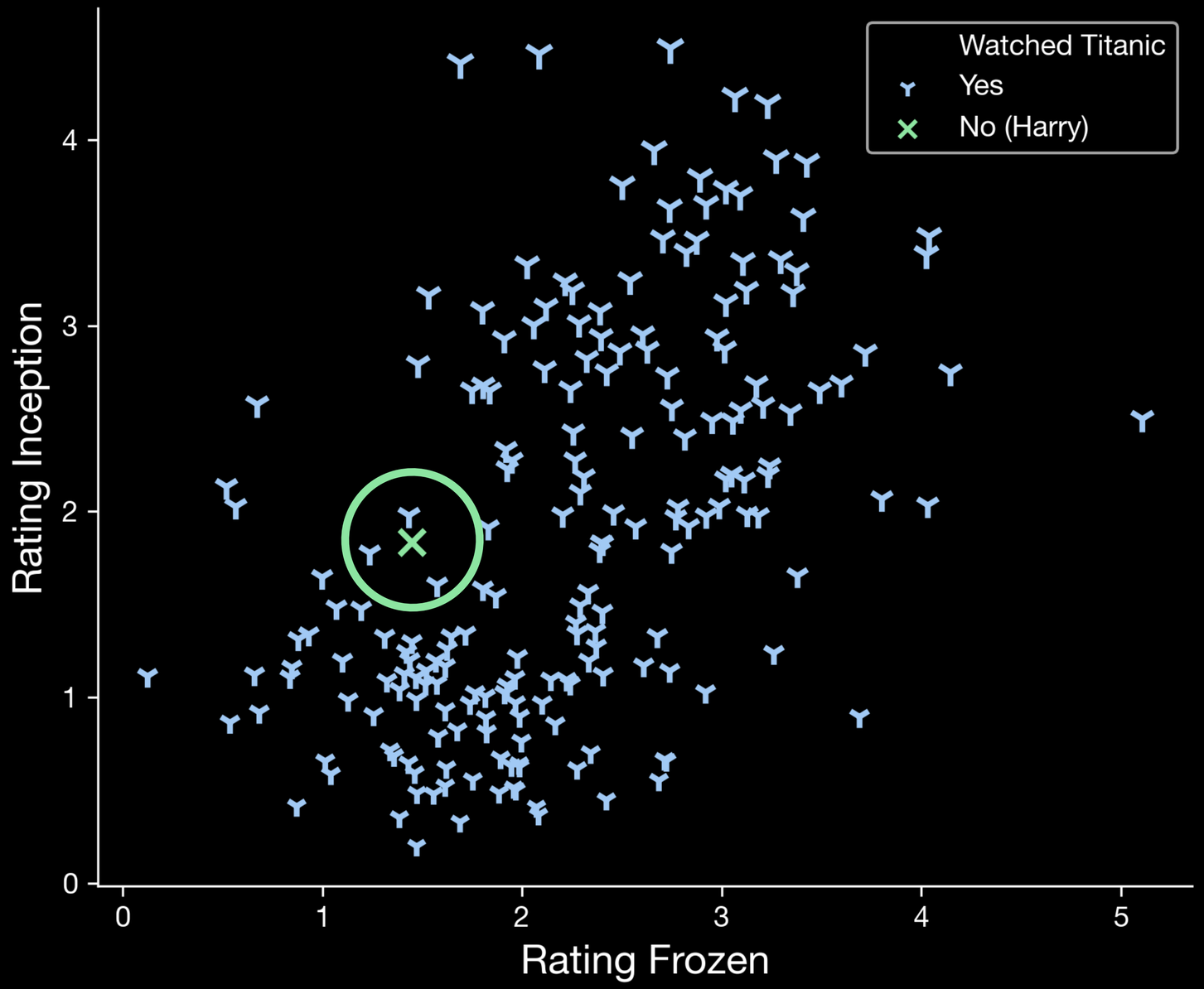
KNN - Movies

Standaard *machine learning* classificatie-algoritme

Rating (Froze)	Rating (Incep)	Rating (Titan)	Distance (to Harry)
1.434	1.979	2.014	0.1750
1.238	1.777	2.629	0.1824
1.576	1.606	1.382	0.2911

mean

2.008



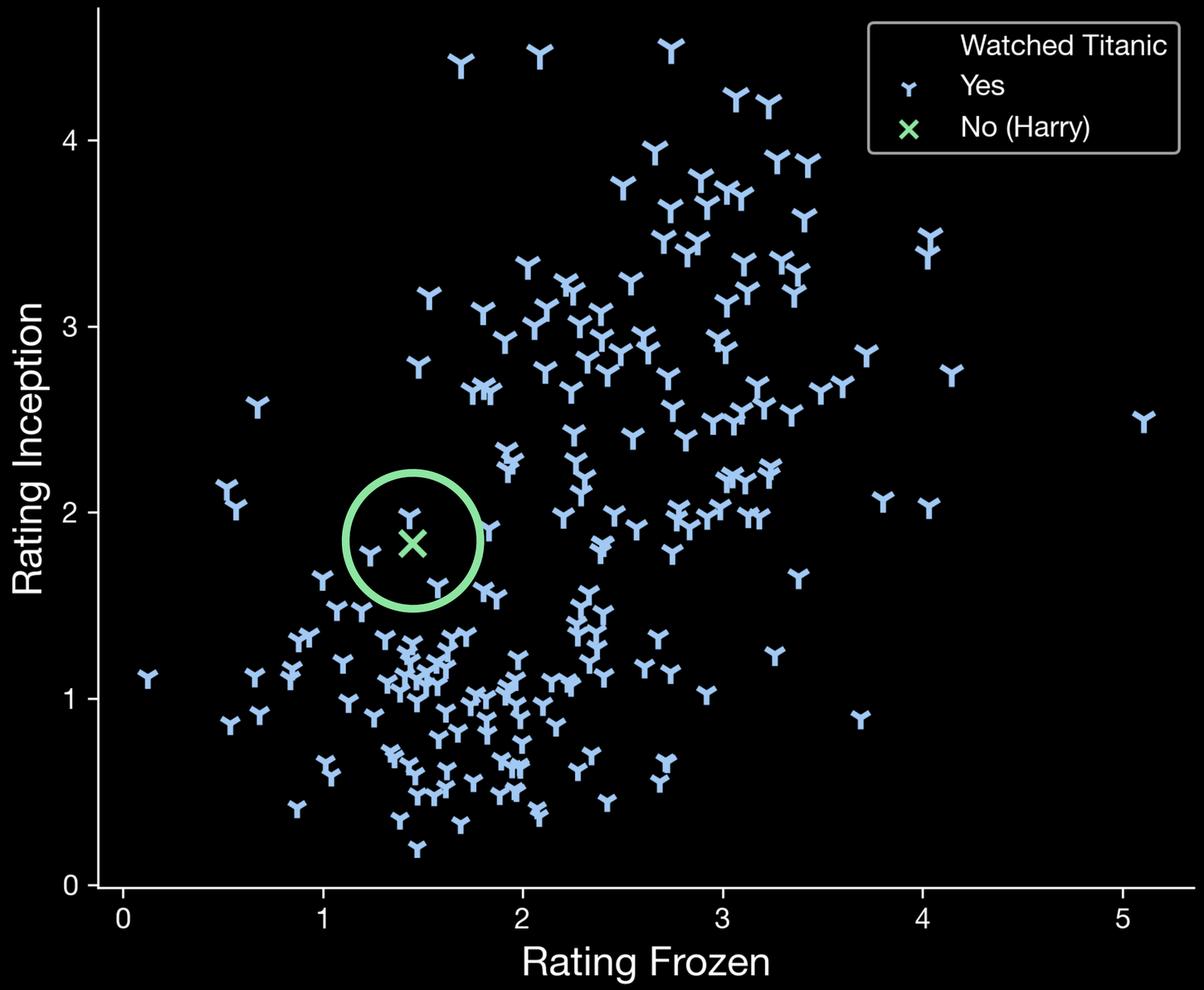
KNN - Movies

Standaard *machine learning* classificatie-algoritme

Rating (Froze)	Rating (Incep)	Rating (Titan)	Distance (to Harry)
1.434	1.979	2.014	0.1750
1.238	1.777	2.629	0.1824
1.576	1.606	1.382	0.2911

weighted mean

?



KNN - Movies

Standaard *machine learning* classificatie-algoritme

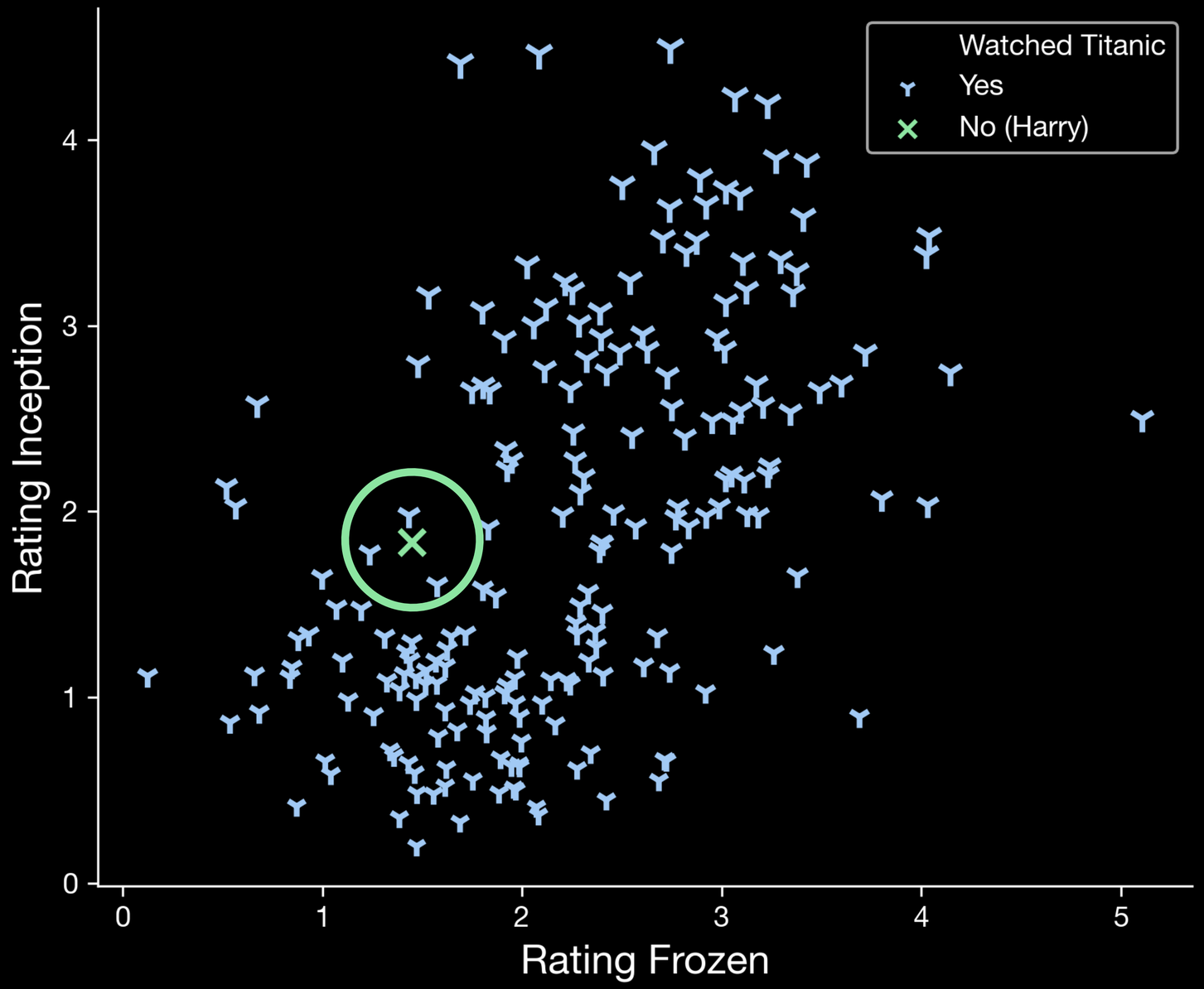
```
similarity = 1 / distance
```

```
Rating Similarity

2.014      5.7142
2.629      5.4824
1.382      3.4352

weighted mean

2.0960
```



KNN - Movies

Standaard *machine learning* classificatie-algoritme

Rating Similarity

2.014	5.7142
2.629	5.4824
1.382	3.4352

↓

2.0960

↓ treshold (> 3.5)

Geen aanbeveling

KNN

K?

(hoe groot kiezen we K)

NN?

(Wat is “near” in “nearest neighbor”)

KNN

K?

(hoe groot kiezen we K)

NN?

(Wat is “near” in “nearest neighbor”)

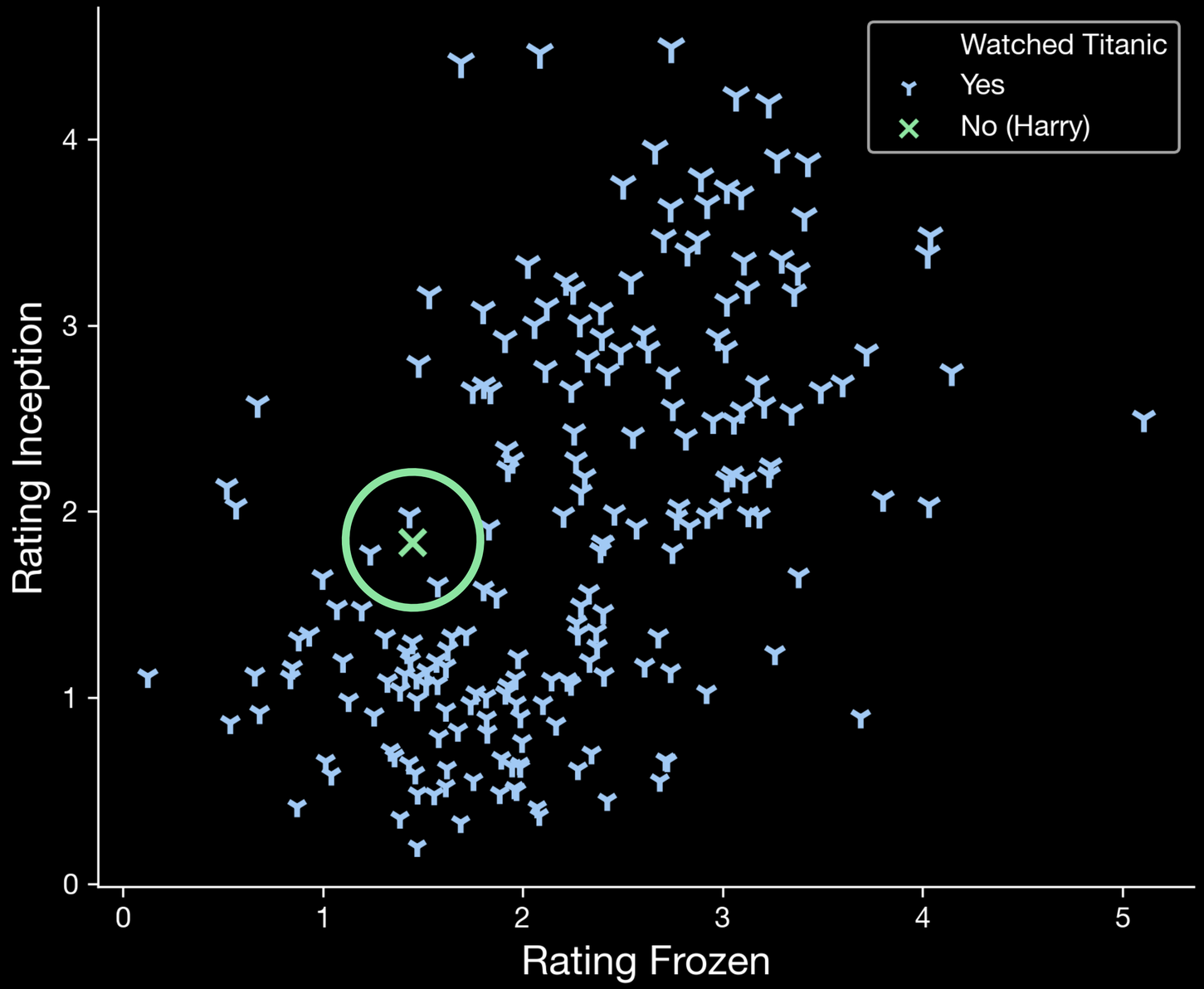
$$Cos(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u) \cdot (r_{v,i} - \mu_v)}{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u)^2 \cdot \sum_{i \in I_u \cap I_v} (r_{v,i} - \mu_v)^2}$$

Euclidian distance

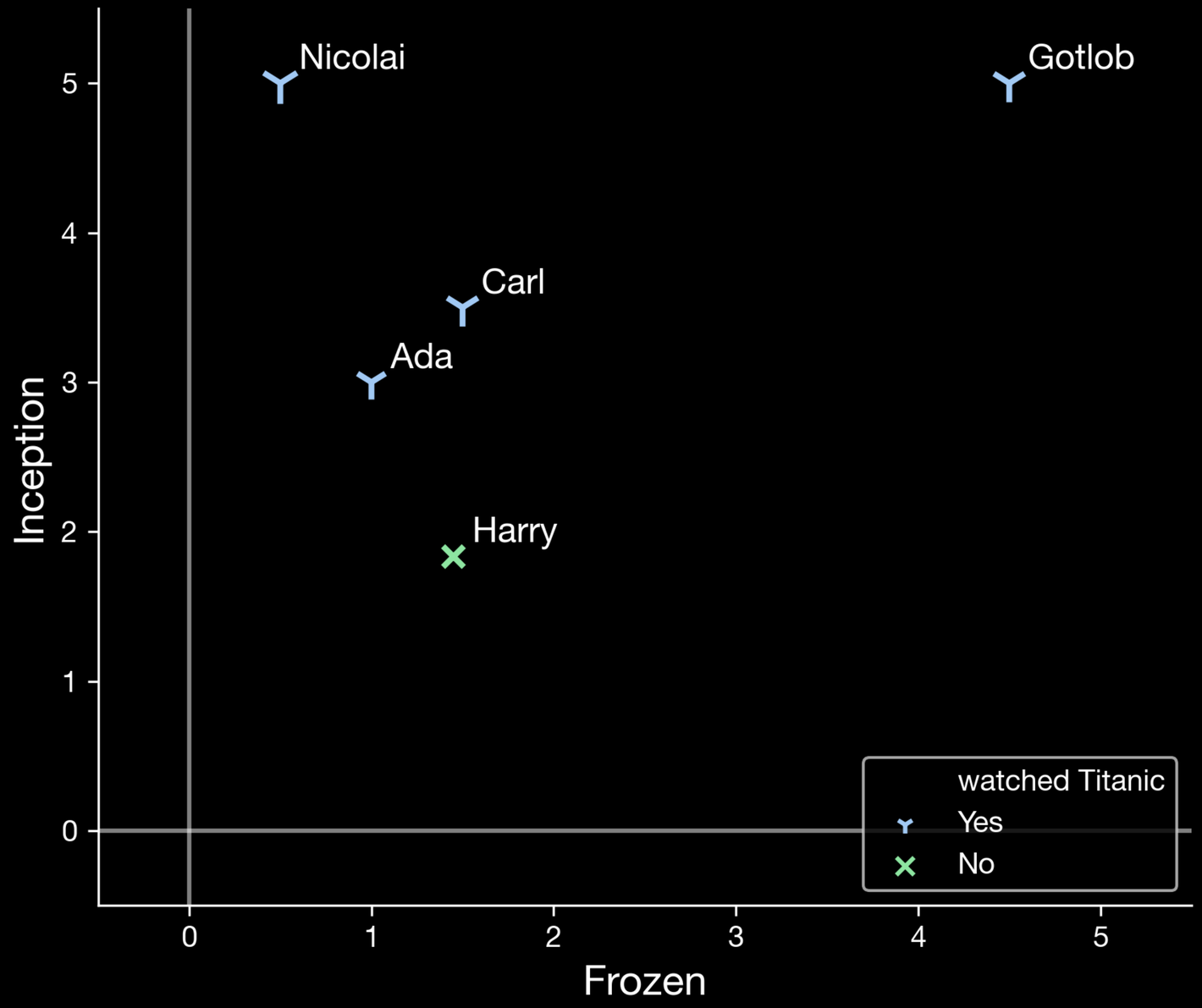
$$d(u,v) = \sqrt{\sum_{i \in I_u \cap I_v} (r_{i,u} - r_{i,v})^2}$$

KNN - Movies

Similarity



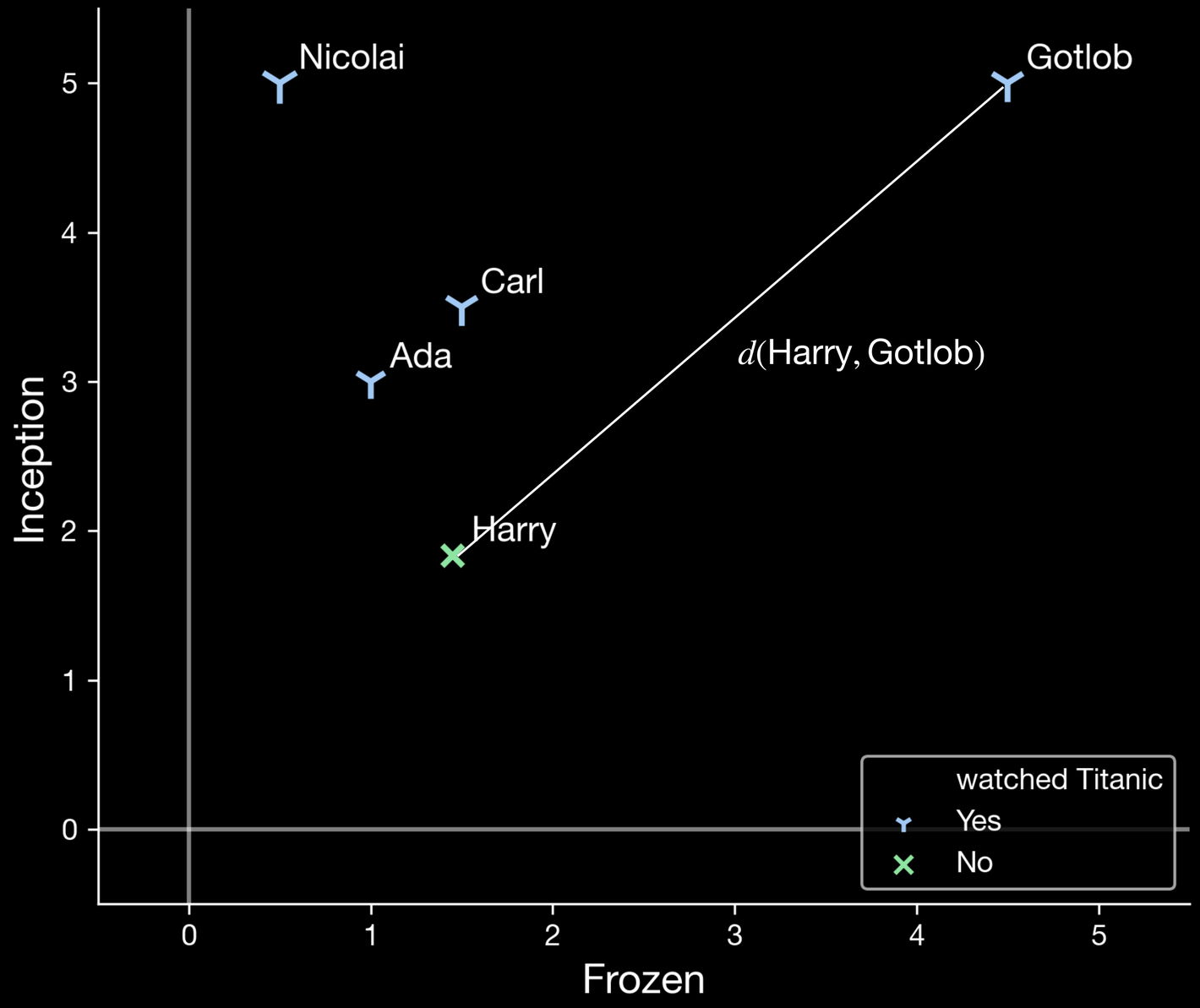
KNN - Movies


Nicolai
Gotlob
Carl
Ada
Harry
Inception
Frozen
0
1
2
3
4
5
watched Titanic
Yes
No

KNN - Movies

Euclidian distance

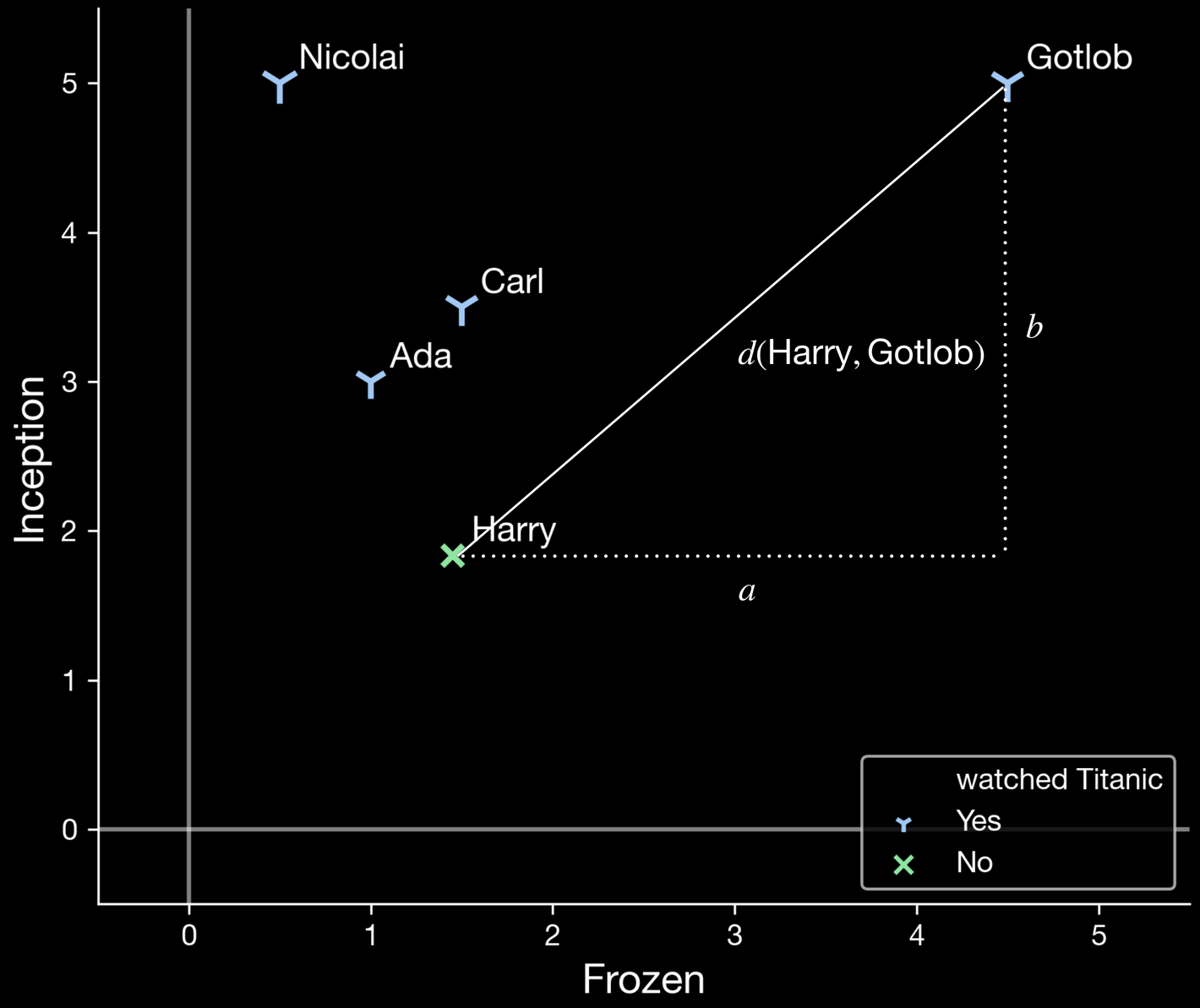
$d(\text{Harry, Gotlob}) = ?$

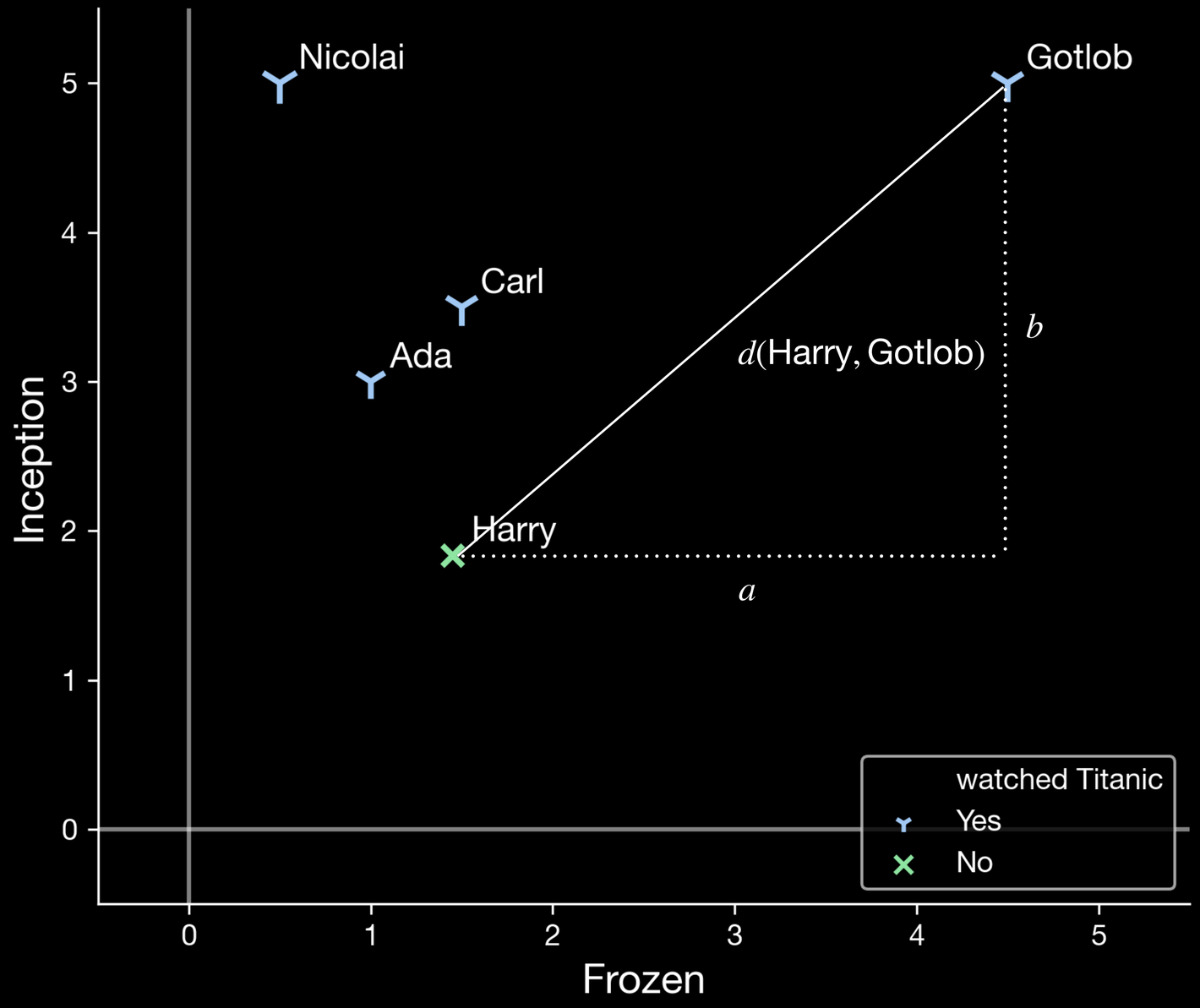


KNN - Movies

Euclidian distance

$d(\text{Harry, Gotlob}) = ?$

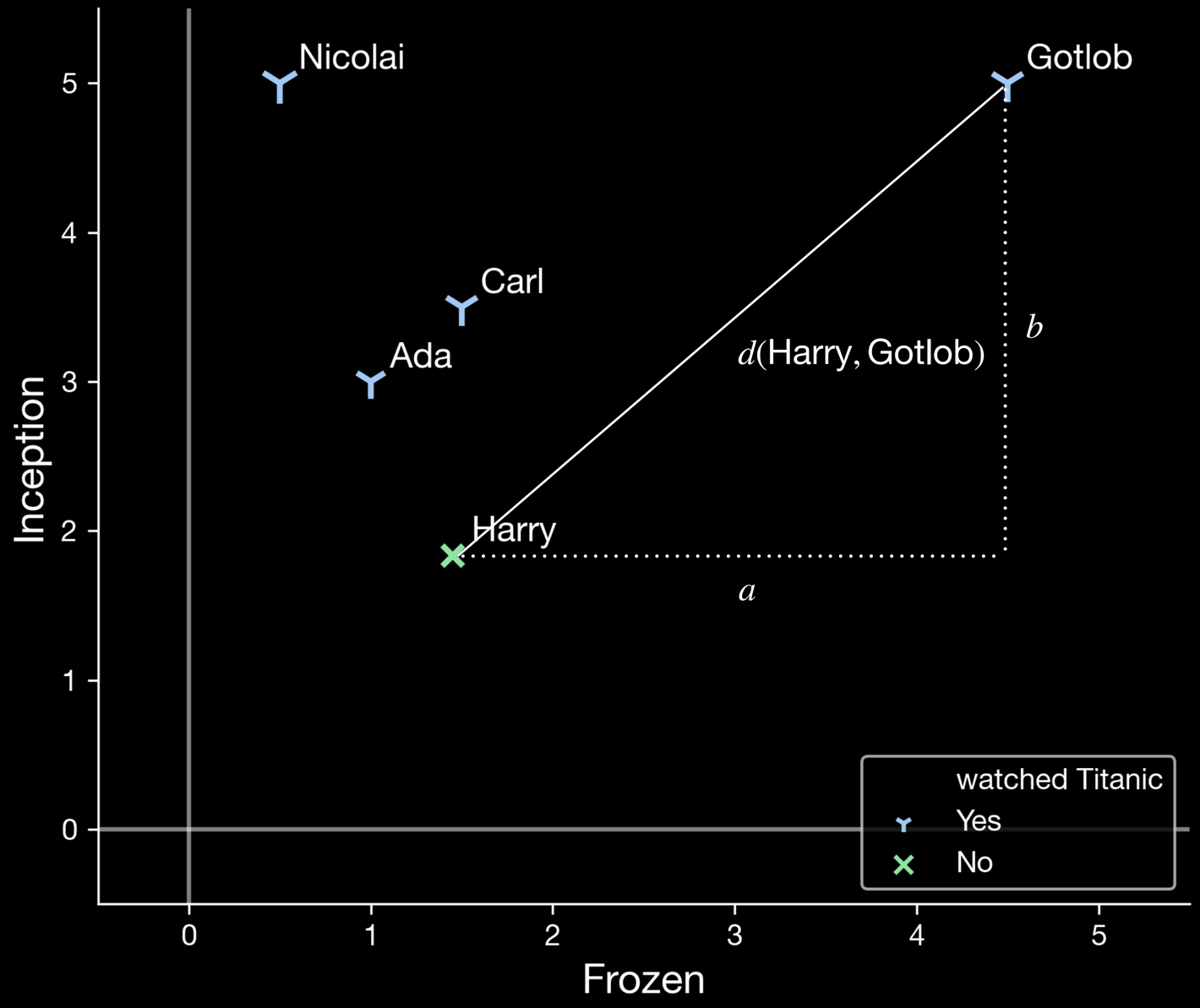




KNN - Movies

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$



KNN - Movies

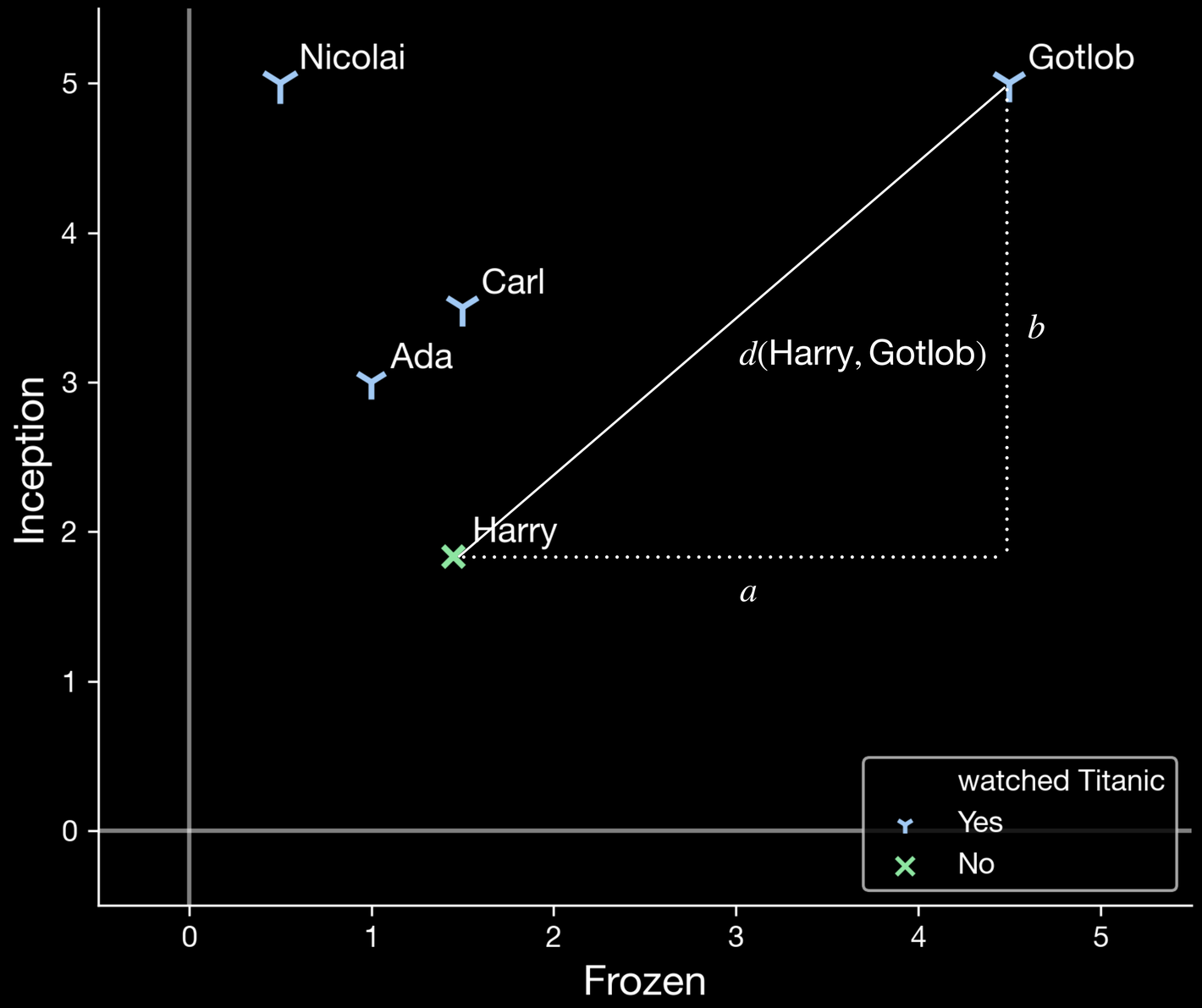
Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$

$$a = r_{frozen,gotlob} - r_{frozen,harry}$$

$$b = r_{inception,gotlob} - r_{inception,harry}$$

	Gotlob	Harry
Frozen	4.5	1.45
Incept	5.0	1.83



KNN - Movies

Euclidian distance

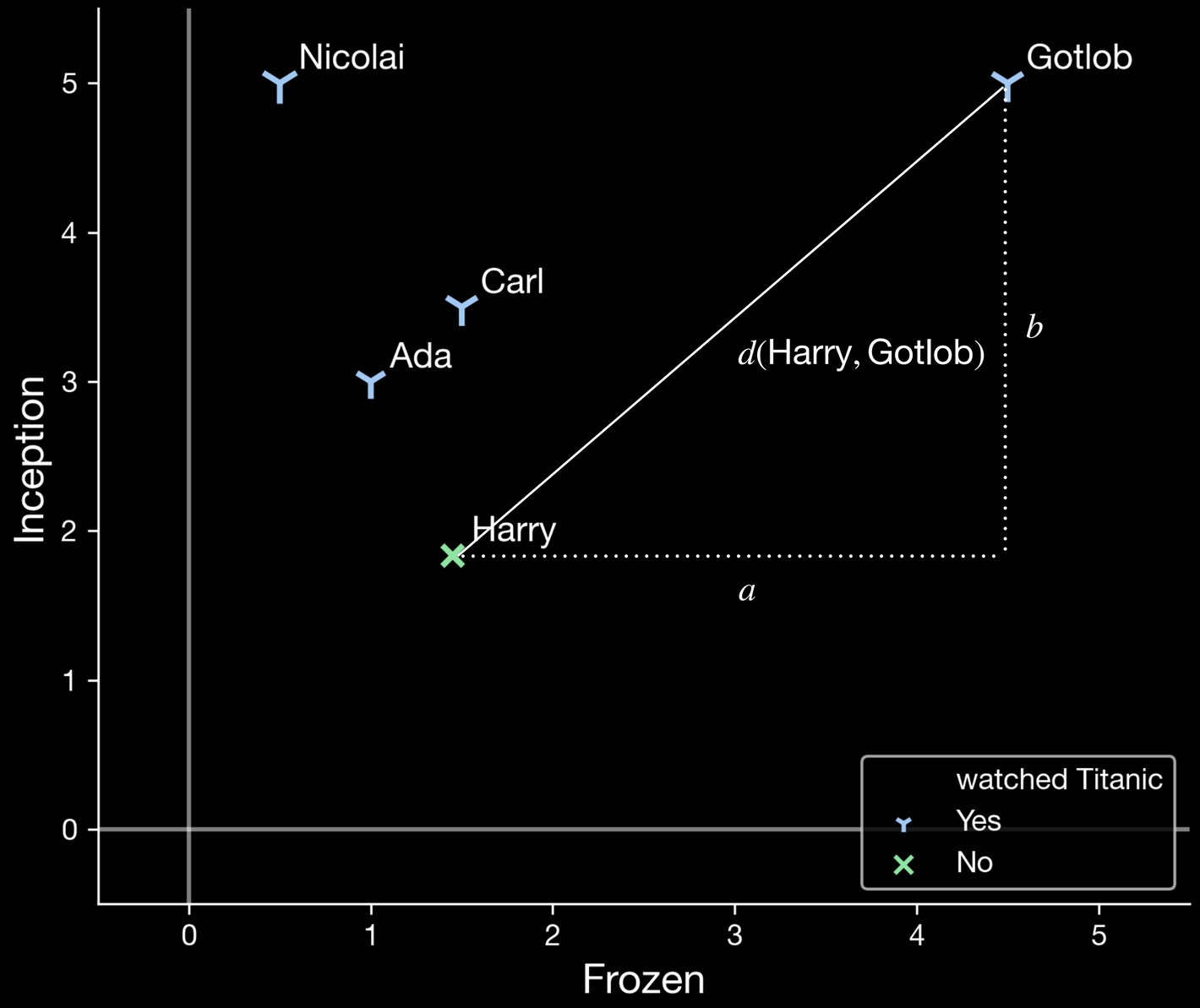
$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$

$$a = r_{frozen,gotlob} - r_{frozen,harry}$$

$$b = r_{inception,gotlob} - r_{inception,harry}$$

```
        Gotlob  Harry
Frozen  4.5 - 1.45 = 3.05

Incept  5.0 - 1.83 = 3.17
```



KNN - Movies

Euclidian distance

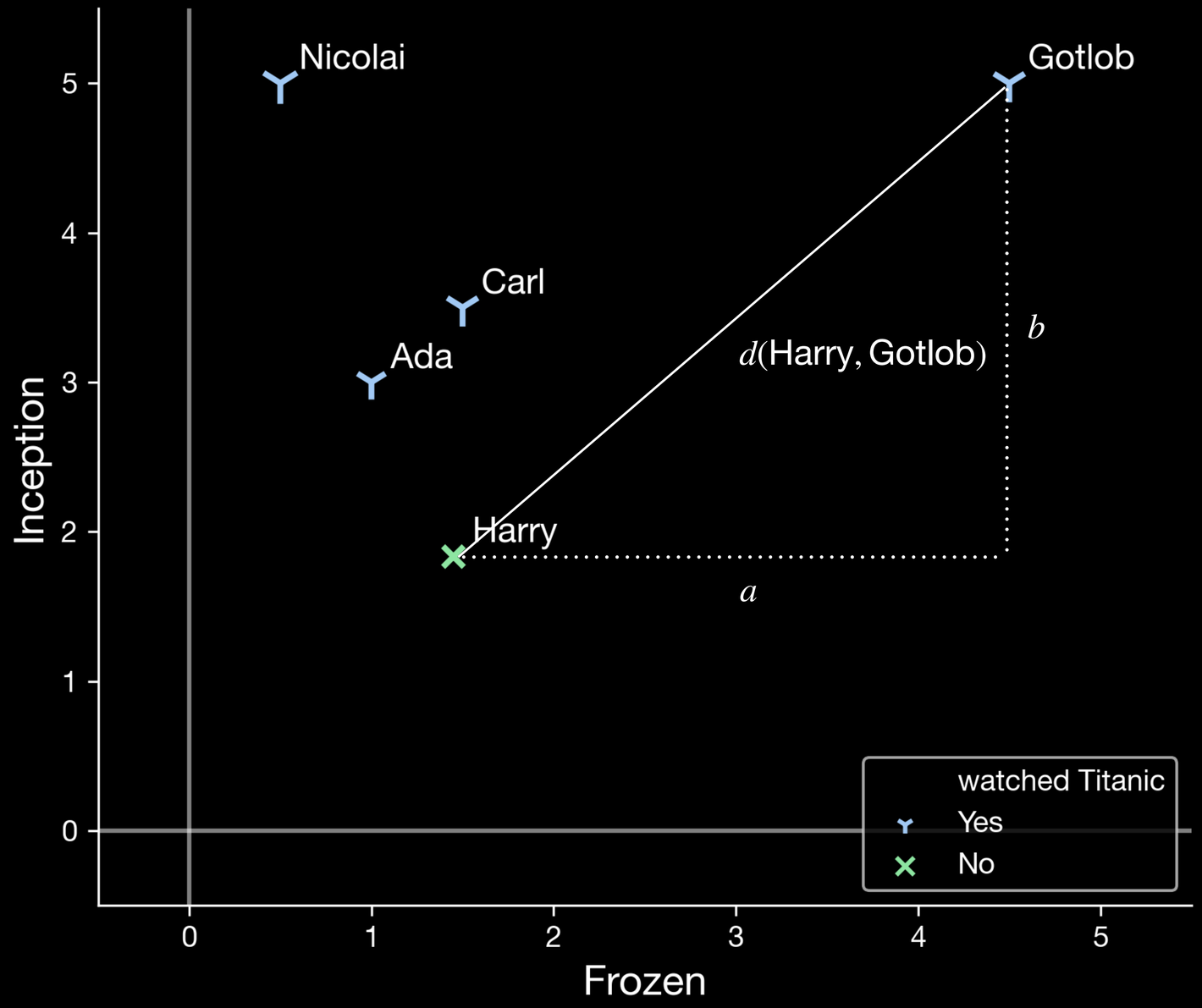
$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$

$$a = r_{frozen,gotlob} - r_{frozen,harry}$$

$$b = r_{inception,gotlob} - r_{inception,harry}$$

```
        Gotlob  Harry
Frozen  4.5   - 1.45   3.05**2 =  9.30

Incept  5.0   - 1.83   3.17**2 = 10.02
```



KNN - Movies

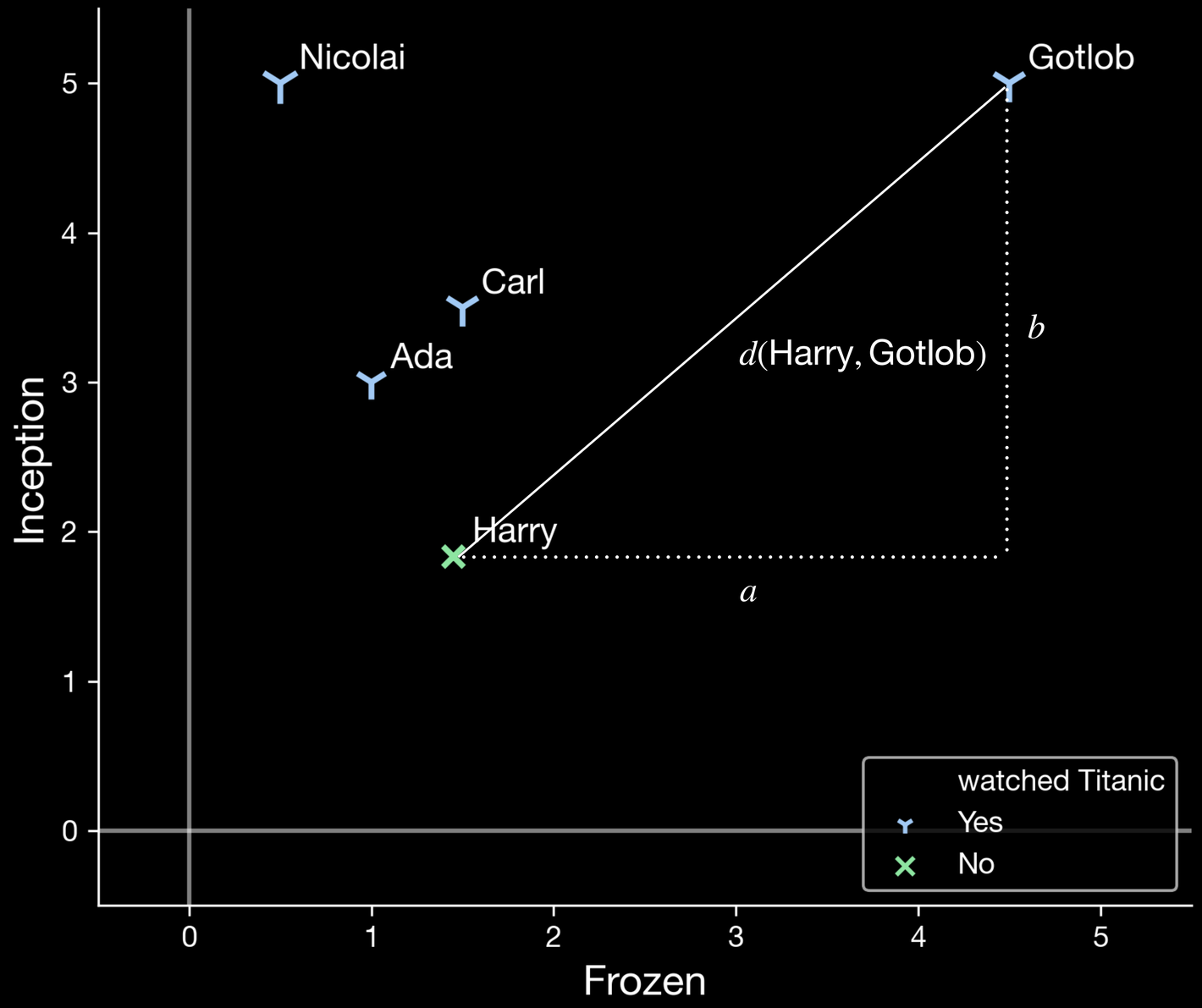
Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$

$$a = r_{frozen,gotlob} - r_{frozen,harry}$$

$$b = r_{inception,gotlob} - r_{inception,harry}$$

```
    Gotlob  Harry
Frozen 4.5 - 1.45   3.05**2 =  9.30
                                +
Incept 5.0 - 1.83   3.17**2 = 10.02
                                =
                              19.33
```



KNN - Movies

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$

$$a = r_{frozen,gotlob} - r_{frozen,harry}$$

$$b = r_{inception,gotlob} - r_{inception,harry}$$

```
    Gotlob  Harry
Frozen 4.5 - 1.45   3.05**2 =  9.30
                              +
Incept 5.0 - 1.83   3.17**2 = 10.02

                          sqrt(19.33)
                                =
                              4.4
```

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{a^2 + b^2}$$

$$a = r_{frozen,gotlob} - r_{frozen,harry}$$

$$b = r_{inception,gotlob} - r_{inception,harry}$$

$$d(\text{Harry, Gotlob}) = \sqrt{(r_{frozen,gotlob} - r_{frozen,harry})^2 + (r_{inception,gotlob} - r_{inception,harry})^2}$$

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{(r_{frozen,gotlob} - r_{frozen,harry})^2 + (r_{inception,gotlob} - r_{inception,harry})^2 + (r_{interstellar,gotlob} - r_{interstellar,harry})^2 + \cdots}$$

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{(r_{frozen,gotlob} - r_{frozen,harry})^2 + (r_{inception,gotlob} - r_{inception,harry})^2 + (r_{interstellar,gotlob} - r_{interstellar,harry})^2 + \cdots}$$

$$d(\text{Harry, Gotlob}) = \sqrt{\text{Som van } (r_{i,gotlob} - r_{i,harry})^2 \text{ voor alle films i die zowel Gotlob als Harry gezien hebben}}$$

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{(r_{frozen,gotlob} - r_{frozen,harry})^2 + (r_{inception,gotlob} - r_{inception,harry})^2 + (r_{interstellar,gotlob} - r_{interstellar,harry})^2 + \cdots}$$

$$d(\text{Harry, Gotlob}) = \sqrt{\text{Som van } (r_{i,gotlob} - r_{i,harry})^2 \text{ voor alle films i die zowel Gotlob als Harry gezien hebben}}$$

$$d(\text{Harry, Gotlob}) = \sqrt{\sum_{i \in I_{gotlob} \cap I_{harry}} (r_{i,gotlob} - r_{i,harry})^2}$$

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{(r_{frozen,gotlob} - r_{frozen,harry})^2 + (r_{inception,gotlob} - r_{inception,harry})^2 + (r_{interstellar,gotlob} - r_{interstellar,harry})^2 + \cdots}$$

$$d(\text{Harry, Gotlob}) = \sqrt{\text{Som van } (r_{i,gotlob} - r_{i,harry})^2 \text{ voor alle films i die zowel Gotlob als Harry gezien hebben}}$$

$$d(\text{Harry, Gotlob}) = \sqrt{\sum_{i \in I_{gotlob} \cap I_{harry}} (r_{i,gotlob} - r_{i,harry})^2}$$

```
euclidian_dist = 0
for i in films:
    if i seen by gotlob and harry:
        euclidian_dist += (rating(i, gotlob) - rating(i, harry))**2
```

Euclidian distance

$$d(\text{Harry, Gotlob}) = \sqrt{(r_{frozen,gotlob} - r_{frozen,harry})^2 + (r_{inception,gotlob} - r_{inception,harry})^2 + (r_{interstellar,gotlob} - r_{interstellar,harry})^2 + \dots}$$

$$d(\text{Harry, Gotlob}) = \sqrt{\sum_{i \in I_{gotlob} \cap I_{harry}} (r_{i,gotlob} - r_{i,harry})^2}$$

```
euclidian_dist = 0
for i in films:
    if i seen by gotlob and harry:
        euclidian_dist += (rating(i, gotlob) - rating(i, harry))**2
```

```
euclidian_dist = sqrt(((ratings['Harry'] - ratings['Gotlob'])**2).sum())
```

$$d(u,v) = \sqrt{\sum_{i \in I_u \cap I_v} (r_{i,u} - r_{i,v})^2}$$

$$d(user1,user2) = \sqrt{\sum_{film \in Films_{user1\ en\ user2}} (rating_{film,user1} - rating_{film,user2})^2}$$

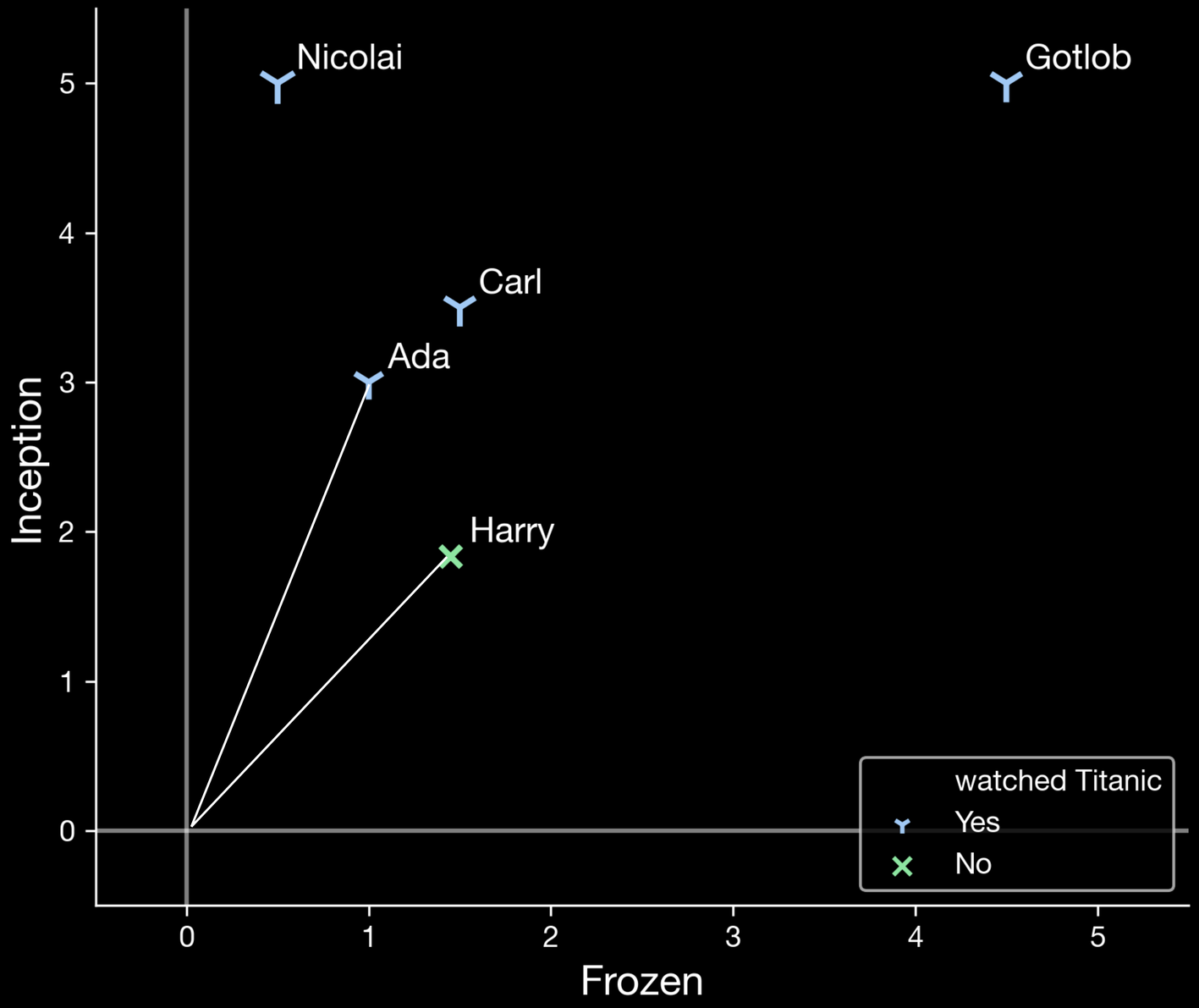
Cosine similarity

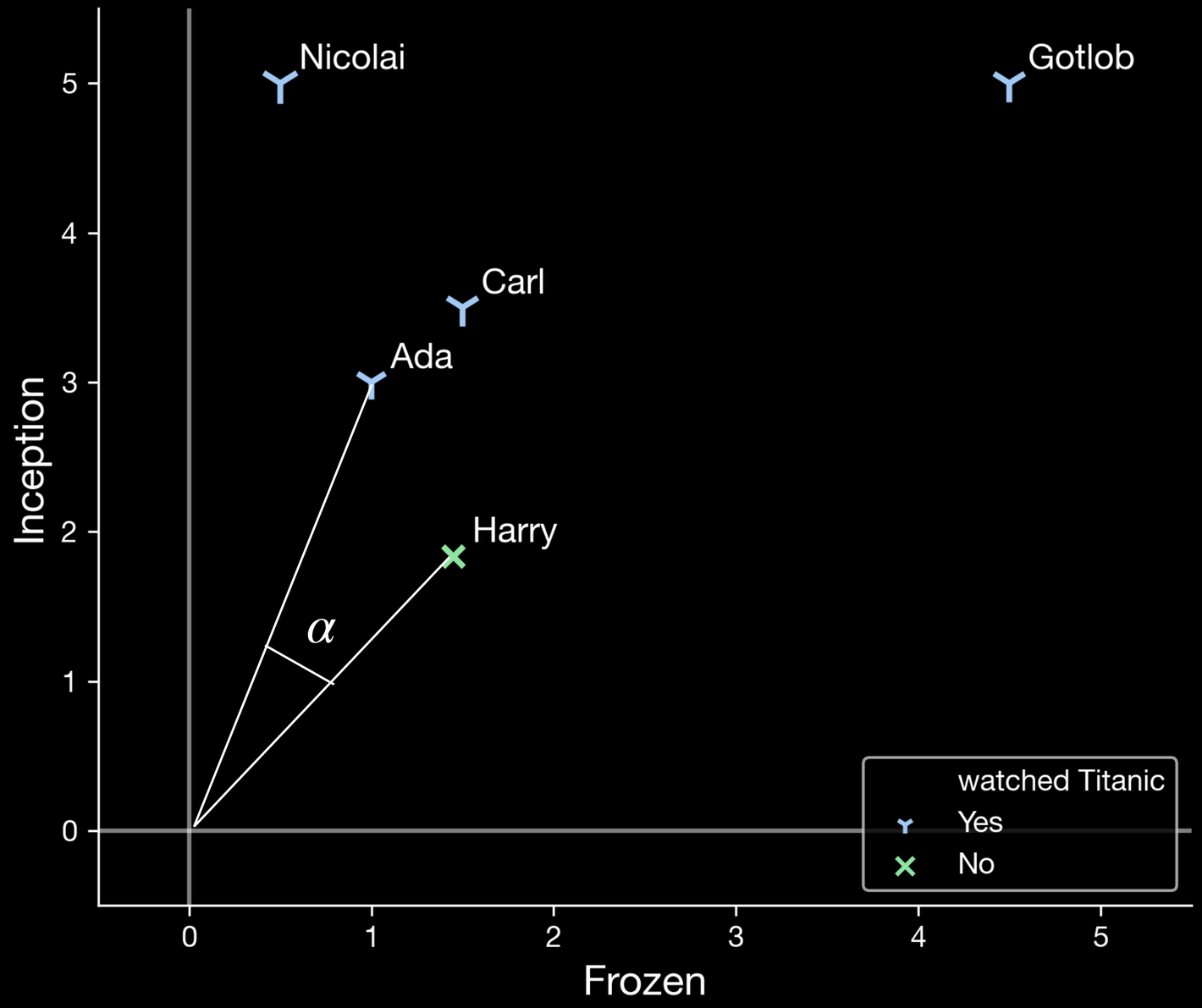
$$Cos(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u) \cdot (r_{v,i} - \mu_v)}{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u)^2 \cdot \sum_{i \in I_u \cap I_v} (r_{v,i} - \mu_v)^2}$$

KNN - Movies

Similarity

$Cos(\text{Harry}, \text{Ada}) = ?$

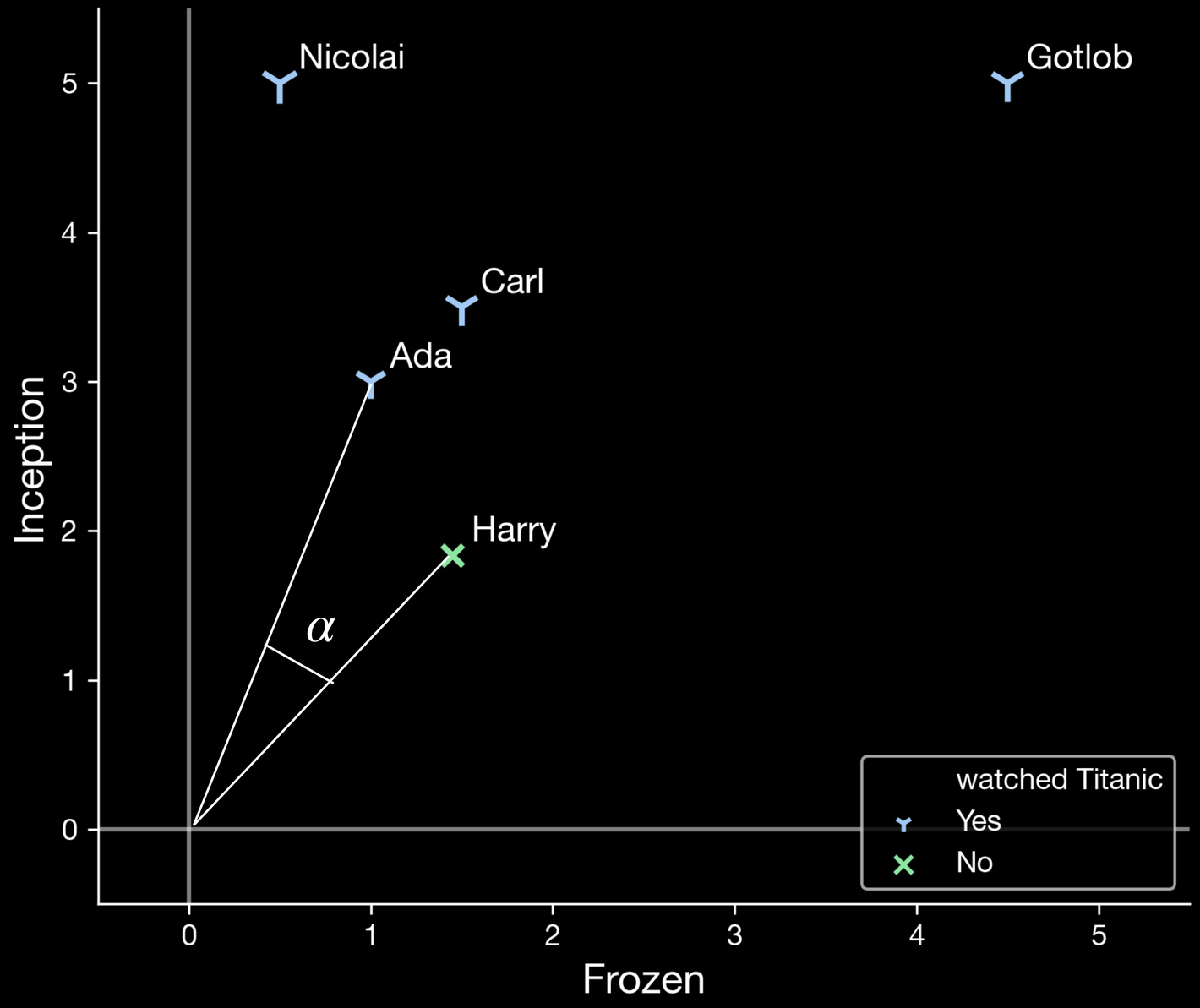




KNN - Movies

Similarity

$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$



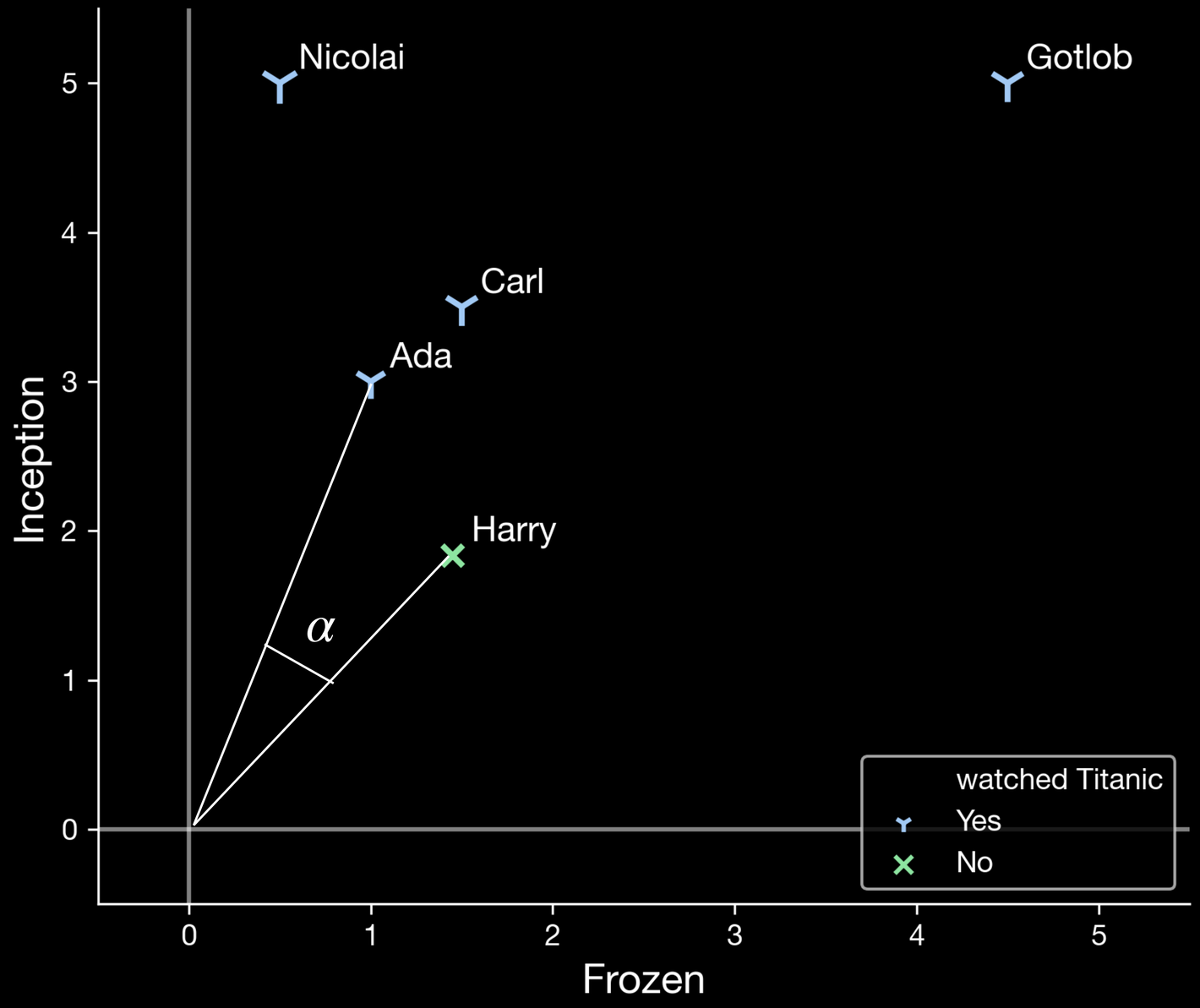
KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

	Harry	Ada
Frozen	1.45	1.00
Incept	1.83	3.00



KNN - Movies

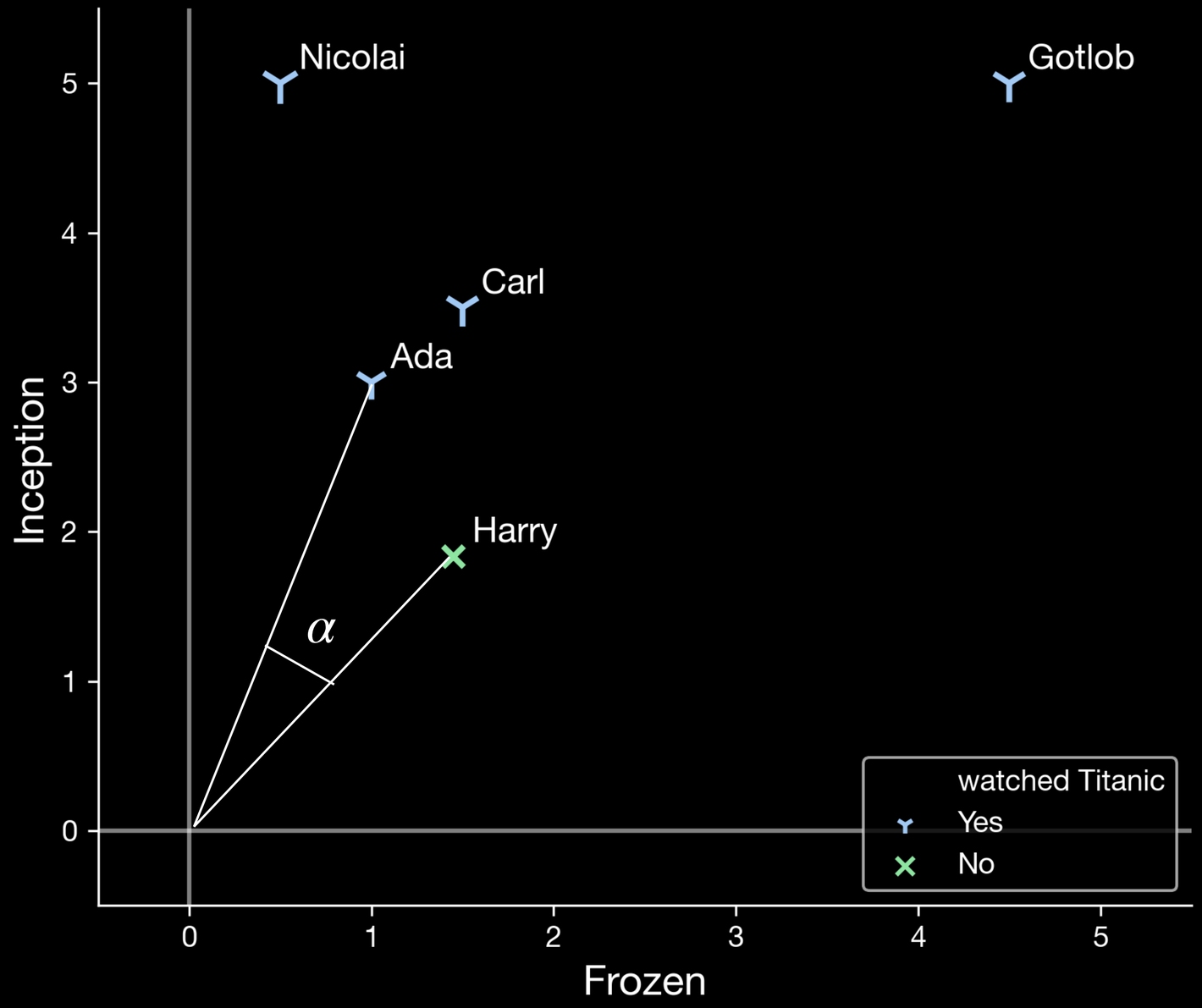
Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

```
        Harry     Ada
Frozen  1.45   * 1.00    = 1.45

Incept  1.83   * 3.00    = 5.50
```



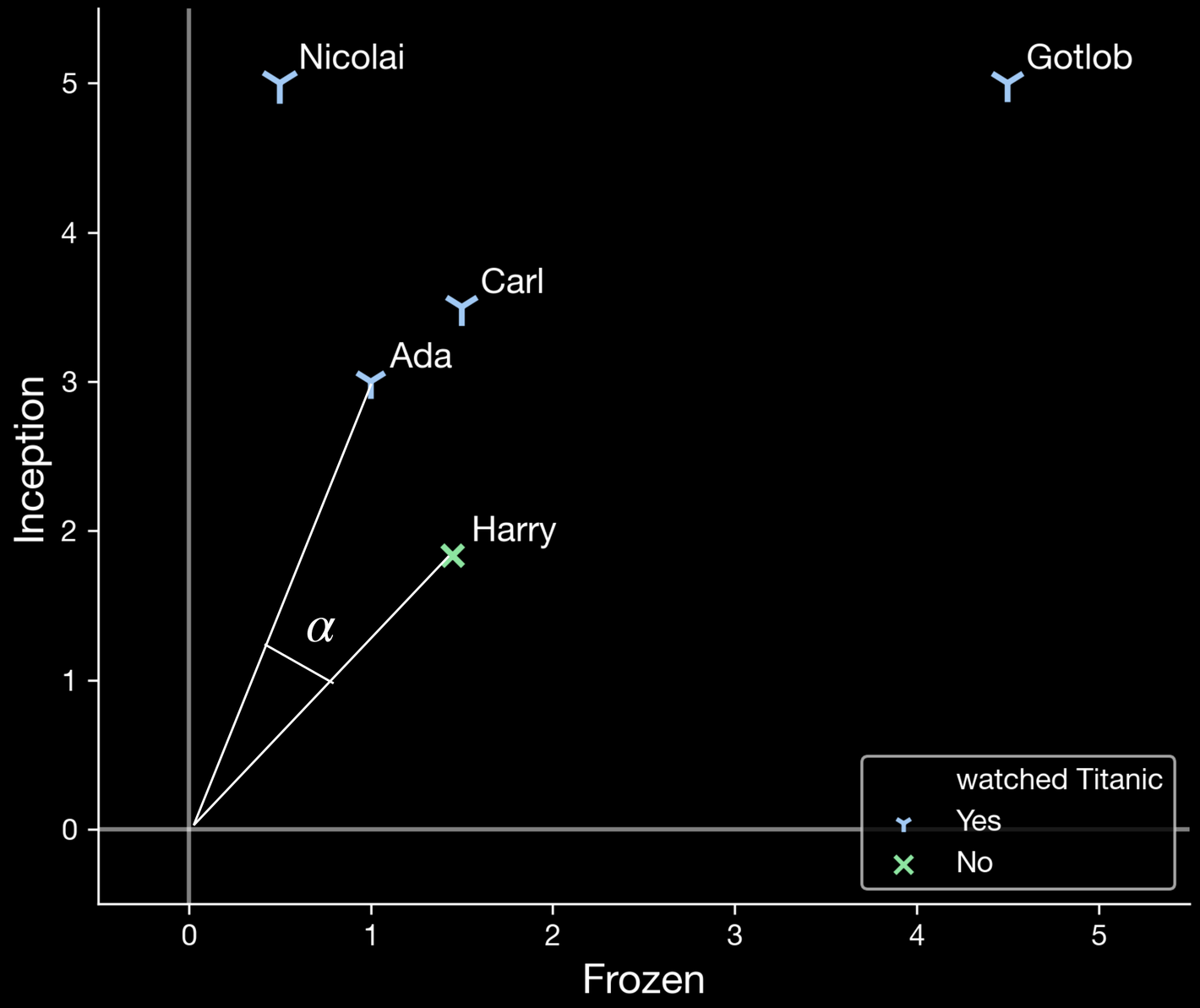
KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

```
       Harry     Ada
Frozen 1.45   * 1.00    = 1.45
                          +
Incept 1.83   * 3.00    = 5.50
                          =
                          6.95
```



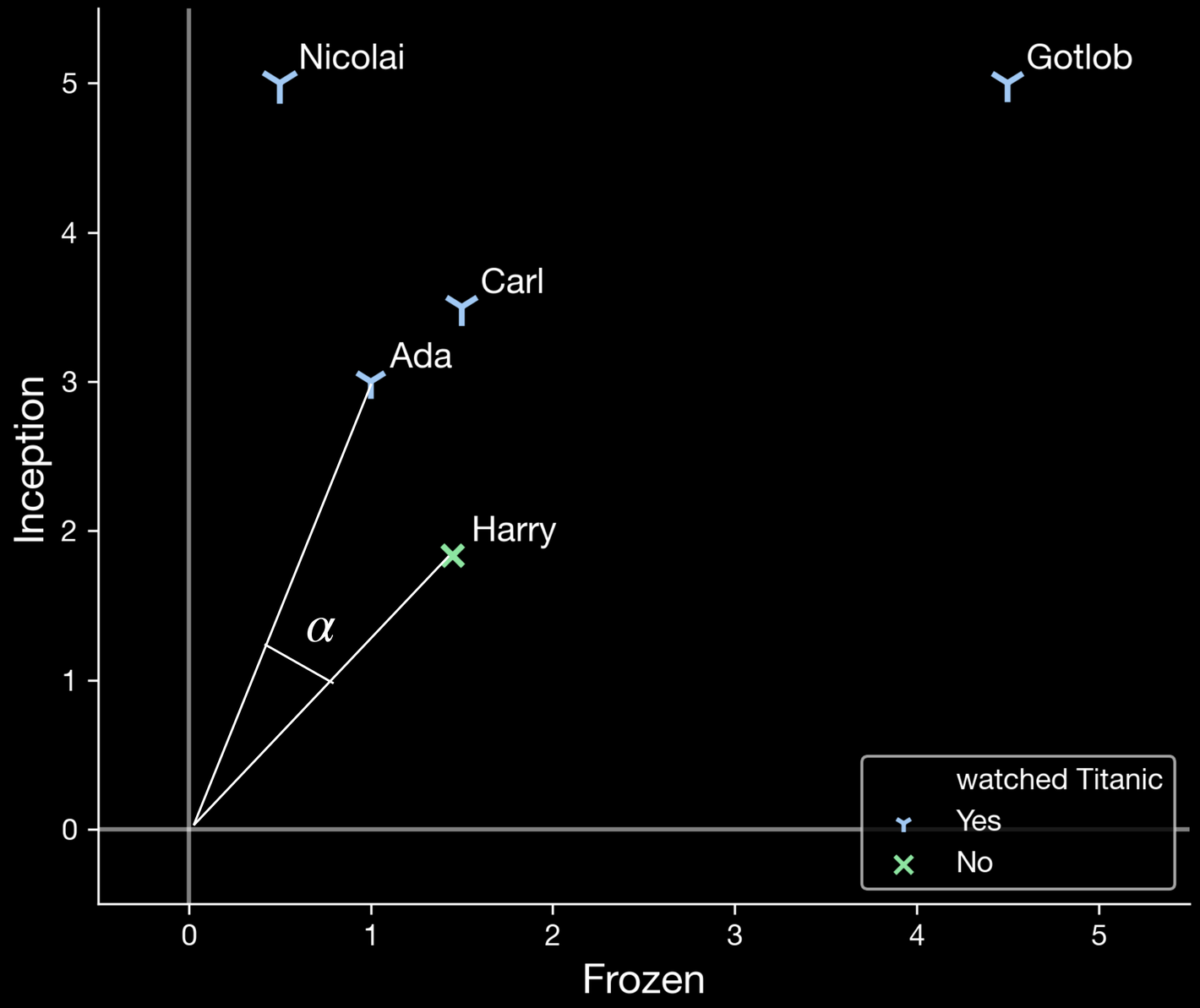
KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

```
        Harry     Ada
Frozen  1.45   * 1.00    = 1.45
                           +
Incept  1.83   * 3.00    = 5.50
                           =
                          6.95
```



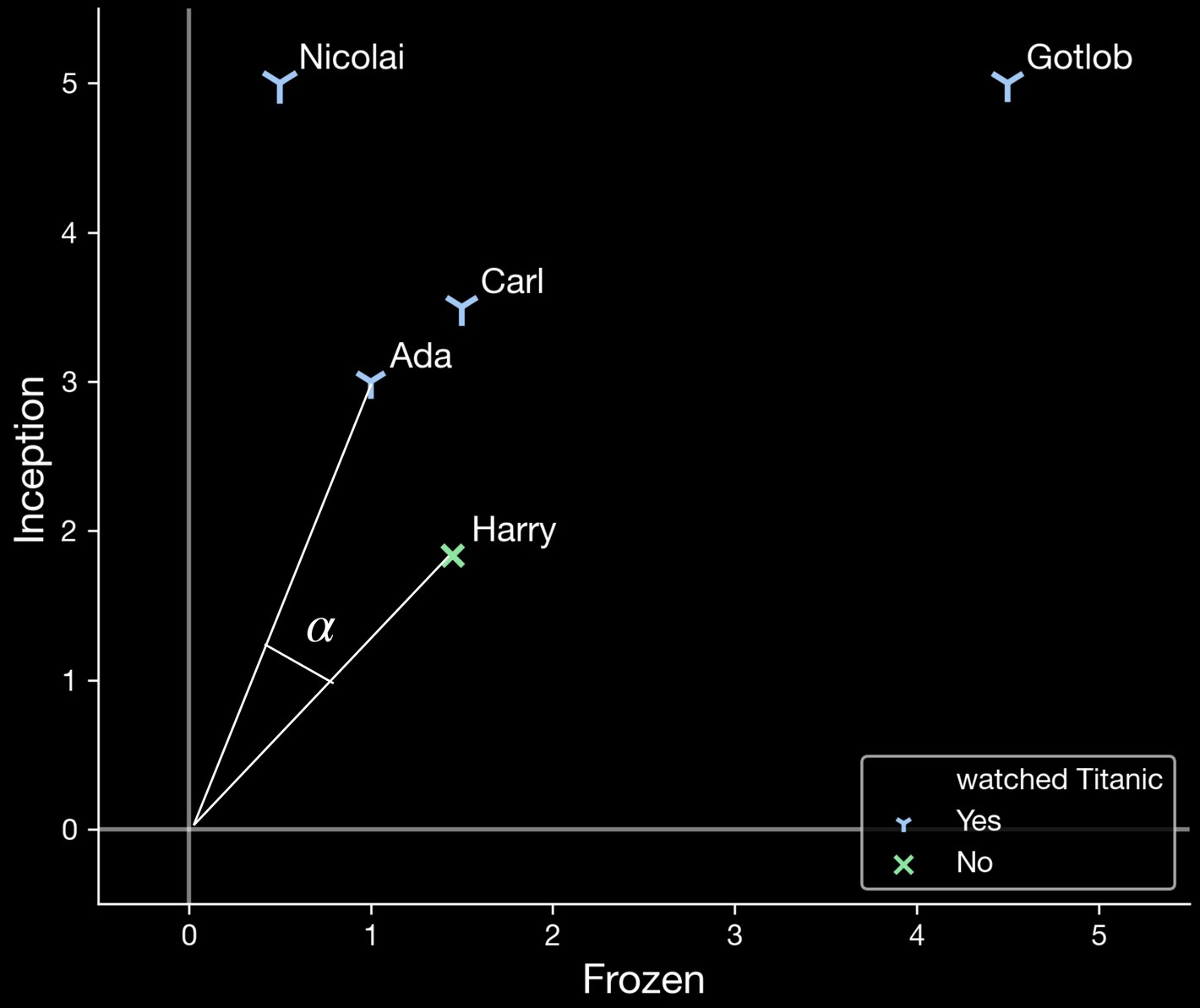
KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

	Harry	Ada	
Frozen	1.45	1.00	1.45
			+
Incept	1.83	3.00	5.50
			=
			6.95



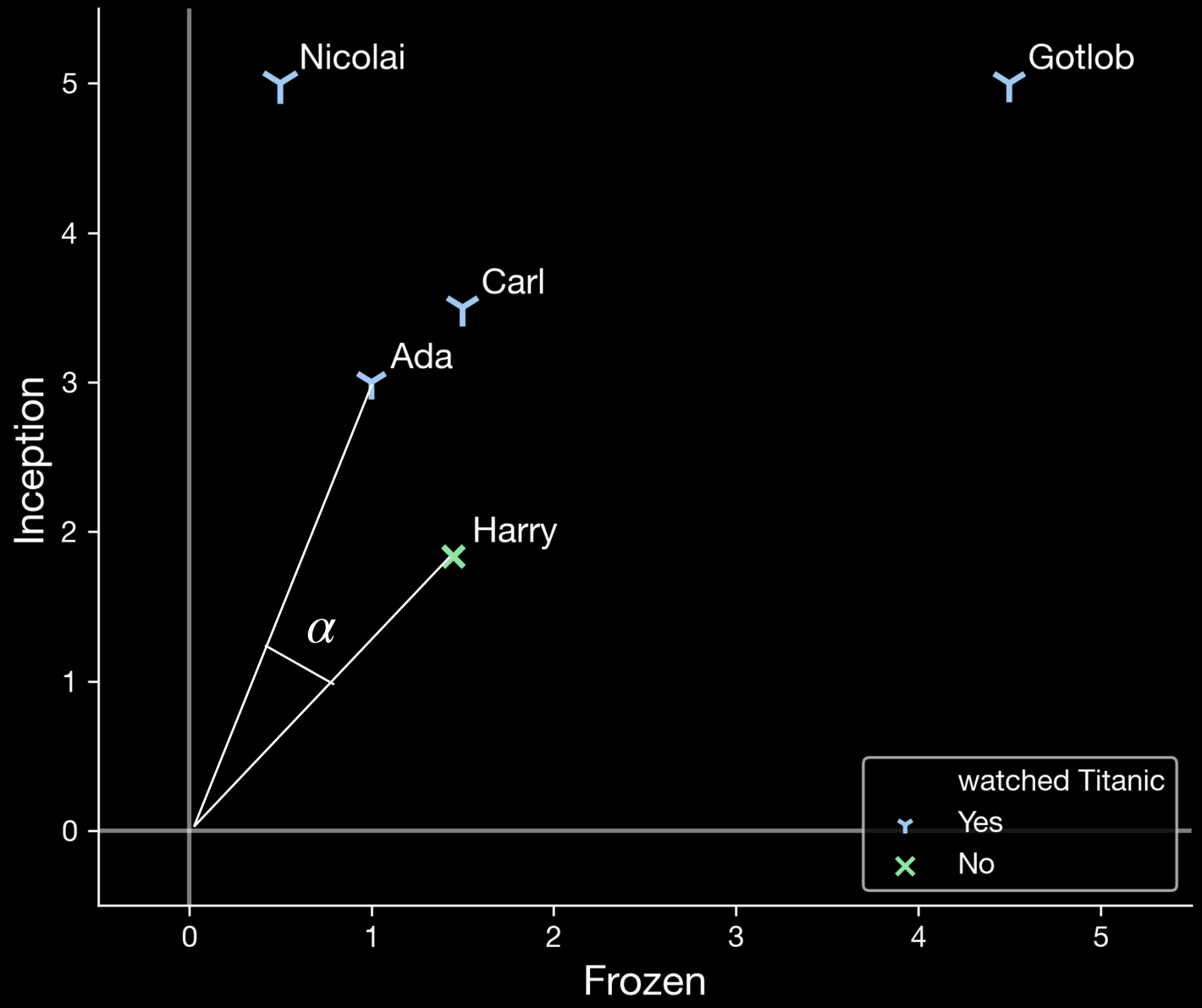
KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

```
        Harry     Ada
Frozen 1.45**2 1.00**2   1.45
                           +
Incept 1.83**2 3.00**2   5.50
                           =
                         6.95
```



KNN - Movies

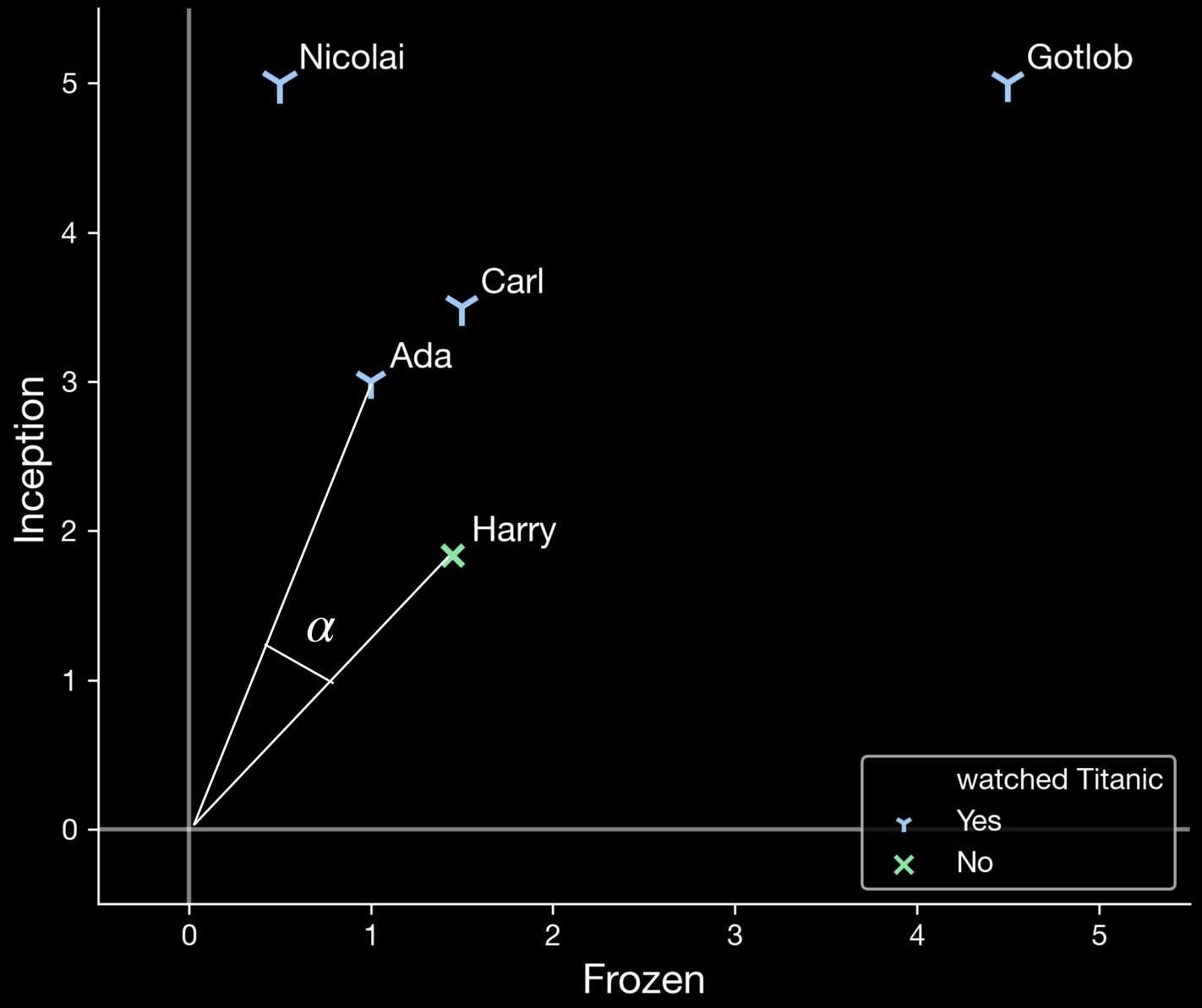
Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

```
        Harry      Ada
Frozen  1.45**2  1.00**2    1.45
           +        +          +
Incept  1.83**2  3.00**2    5.50
                                =
           =        =       6.95

        5.46     10.0
```



KNN - Movies

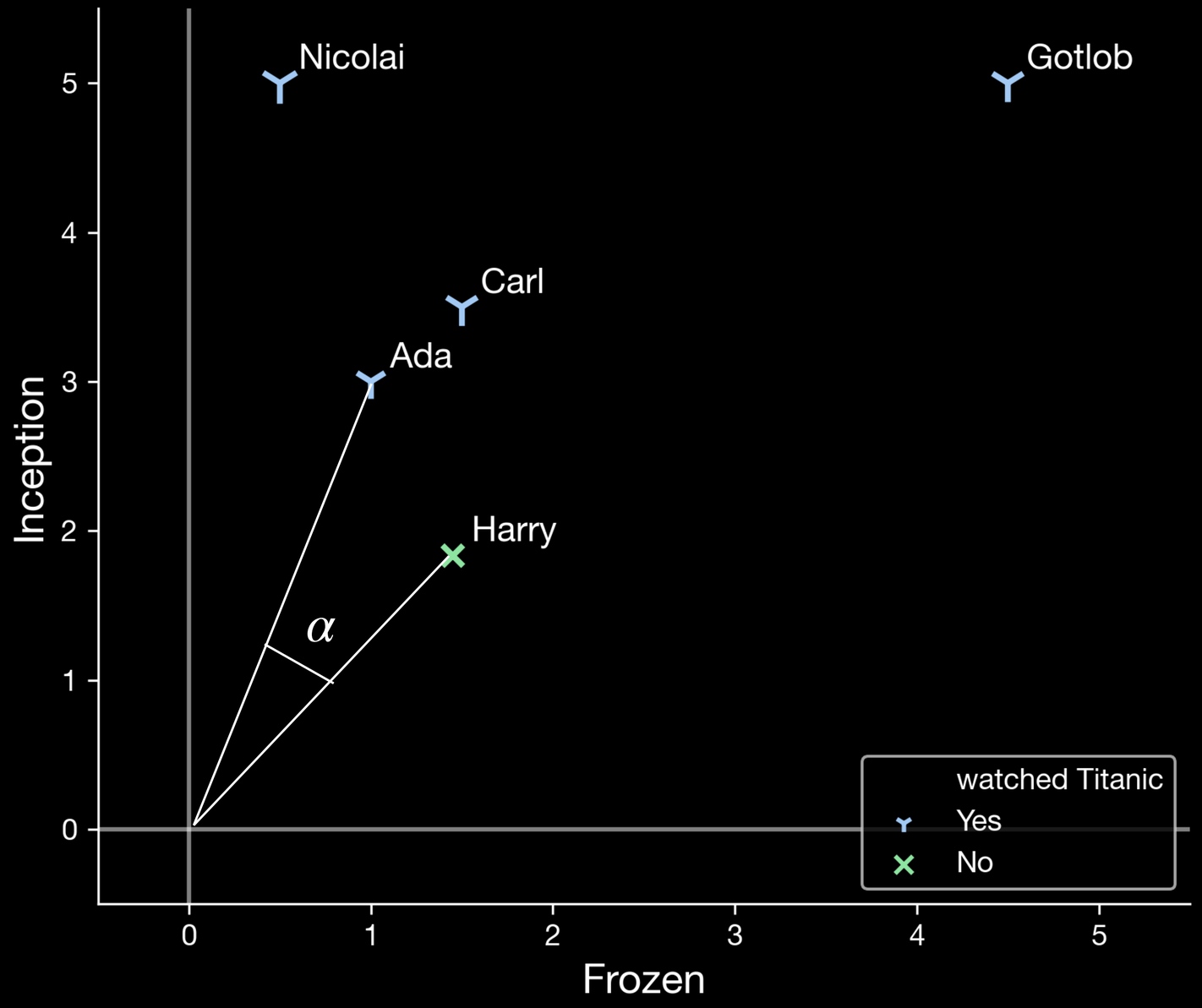
Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

```
        Harry     Ada
Frozen 1.45**2 1.00**2   1.45
          +       +        +
Incept 1.83**2 3.00**2   5.50
                         =
          =       =      6.95

       5.46  * 10.0    = 15.5
```



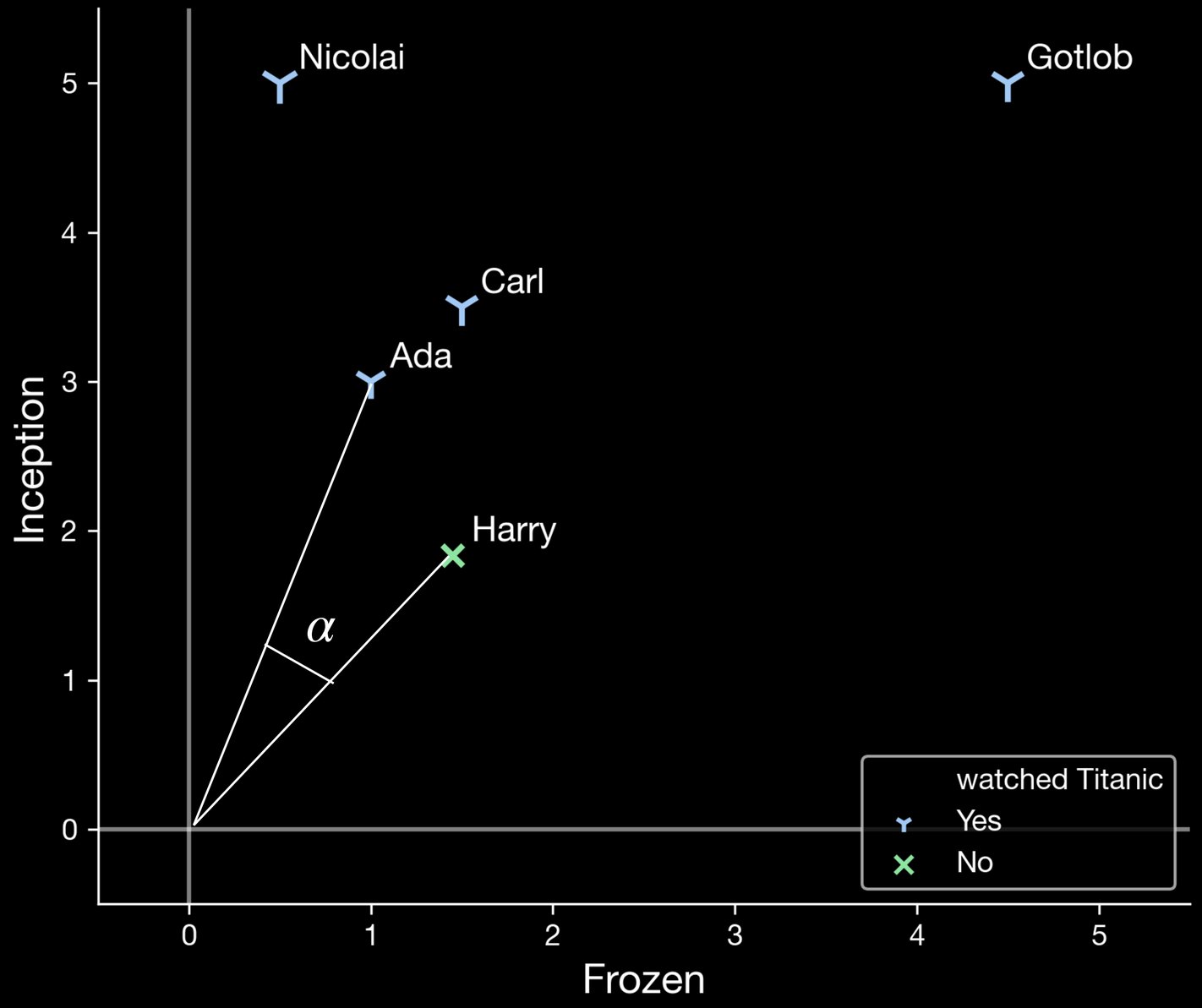
KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

	Harry	Ada	
Frozen	1.45	1.00	1.45
Incept	1.83	3.00	5.50
			6.95
			/
	5.46	10.0	15.5
			=
			0.45



KNN - Movies

Similarity

$$Cos(\text{Harry}, \text{Ada}) = \cos(\alpha)$$

$$cos(\alpha) = \frac{r_{harry,fr} \cdot r_{ada,fr} + r_{harry,in} \cdot r_{ada,in}}{(r^2_{harry,fr} + r^2_{harry,in}) \cdot (r^2_{ada,fr} + r^2_{ada,in})}$$

	Harry	Ada	
Frozen	1.45	1.00	1.45
Incept	1.83	3.00	5.50
			6.95
			/
	5.46	10.0	15.5
			=
			0.45

Euclidian distance

$$Cos(Harry, Ada) = \frac{r_{harry,frozen} \cdot r_{ada,frozen} + r_{harry,incept} \cdot r_{ada,incept} + r_{harry,inters} \cdot r_{ada,inters} + \dots}{(r^2_{harry,frozen} + r^2_{harry,incept} + r^2_{harry,inters} + \dots) \cdot (r^2_{ada,frozen} + r^2_{ada,incept} + r^2_{ada,inters} + \dots)}$$

$$Cos(Harry, Ada) = \frac{\sum_{i \in I_{harry} \cap I_{ada}} r_{harry,i} \cdot r_{ada,i}}{\sum_{i \in I_{harry} \cap I_{ada}} r^2_{harry,i} \cdot \sum_{i \in I_{harry} \cap I_{ada}} r^2_{ada,i}}$$

Euclidian distance

$$Cos(Harry, Ada) = \frac{r_{harry,frozen} \cdot r_{ada,frozen} + r_{harry,incept} \cdot r_{ada,incept} + r_{harry,inters} \cdot r_{ada,inters} + \dots}{(r^2_{harry,frozen} + r^2_{harry,incept} + r^2_{harry,inters} + \dots) \cdot (r^2_{ada,frozen} + r^2_{ada,incept} + r^2_{ada,inters} + \dots)}$$

$$Cos(Harry, Ada) = \frac{\sum_{i \in I_{harry} \cap I_{ada}} r_{harry,i} \cdot r_{ada,i}}{\sum_{i \in I_{harry} \cap I_{ada}} r^2_{harry,i} \cdot \sum_{i \in I_{harry} \cap I_{ada}} r^2_{ada,i}}$$

$$Cos(u, v) = \frac{\sum_{i \in I_u \cap I_v} r_{u,i} \cdot r_{v,i}}{\sum_{i \in I_u \cap I_v} r^2_{u,i} \cdot \sum_{i \in I_u \cap I_v} r^2_{v,i}}$$

Euclidian distance

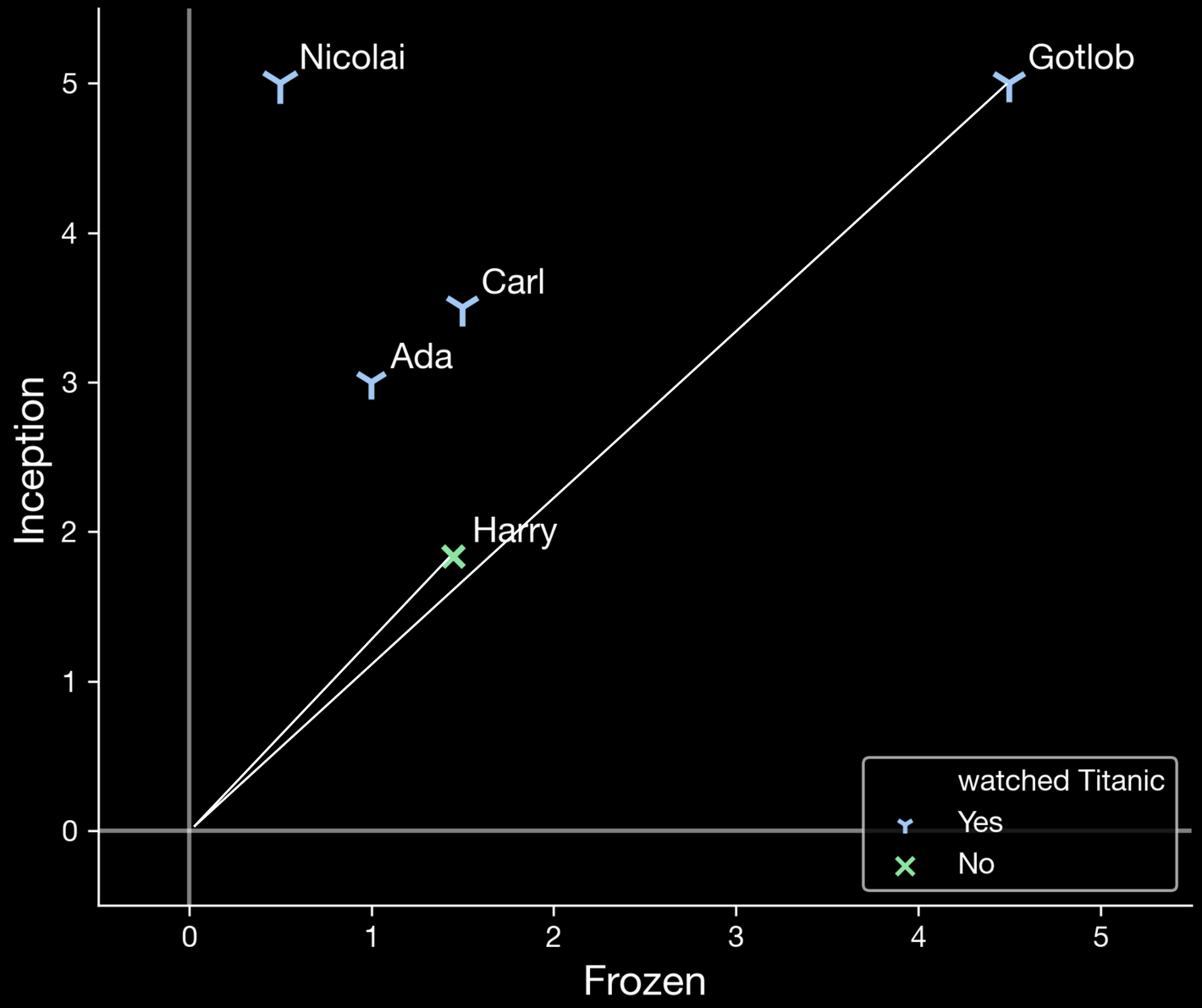
$$Cos(Harry, Ada) = \frac{r_{harry,frozen} \cdot r_{ada,frozen} + r_{harry,incept} \cdot r_{ada,incept} + r_{harry,inters} \cdot r_{ada,inters} + \cdots}{(r^2_{harry,frozen} + r^2_{harry,incept} + r^2_{harry,inters} + \ldots) \cdot (r^2_{ada,frozen} + r^2_{ada,incept} + r^2_{ada,inters} + \ldots)}$$

$$Cos(Harry, Ada) = \frac{\sum_{i \in I_{harry} \cap I_{ada}} r_{harry,i} \cdot r_{ada,i}}{\sum_{i \in I_{harry} \cap I_{ada}} r^2_{harry,i} \cdot \sum_{i \in I_{harry} \cap I_{ada}} r^2_{ada,i}}$$

$$Cos(u, v) = \frac{\sum_{i \in I_u \cap I_v} r_{u,i} \cdot r_{v,i}}{\sum_{i \in I_u \cap I_v} r^2_{u,i} \cdot \sum_{i \in I_u \cap I_v} r^2_{v,i}}$$

$$Cos(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u) \cdot (r_{v,i} - \mu_v)}{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u)^2 \cdot \sum_{i \in I_u \cap I_v} (r_{v,i} - \mu_v)^2}$$

Mean centering

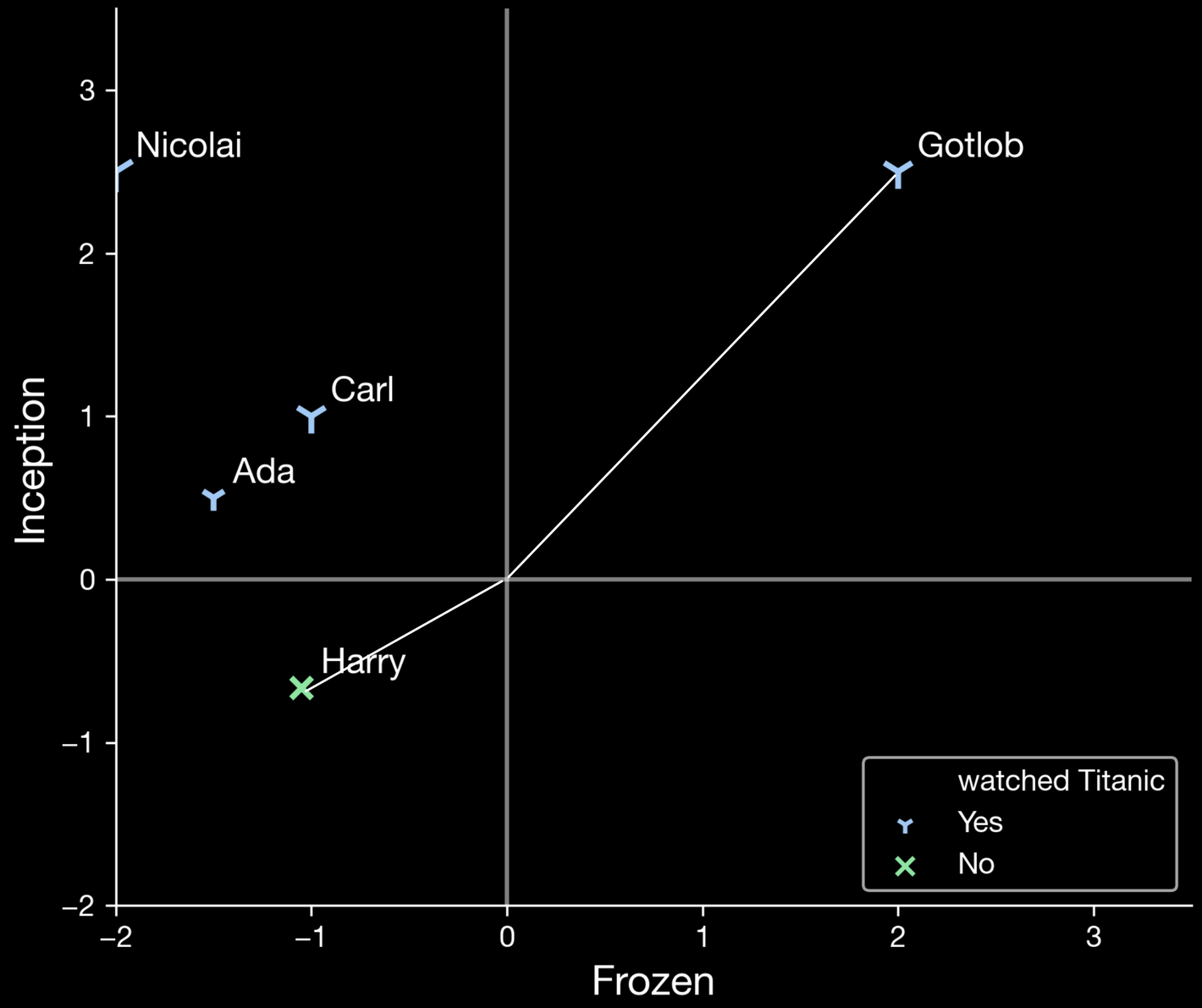


KNN - Movies

Similarity

Sommige mensen ranken altijd positiever dan anderen.

Absolute rating is niet belangrijk, maar afwijking ten opzichte van gemiddeld rating.

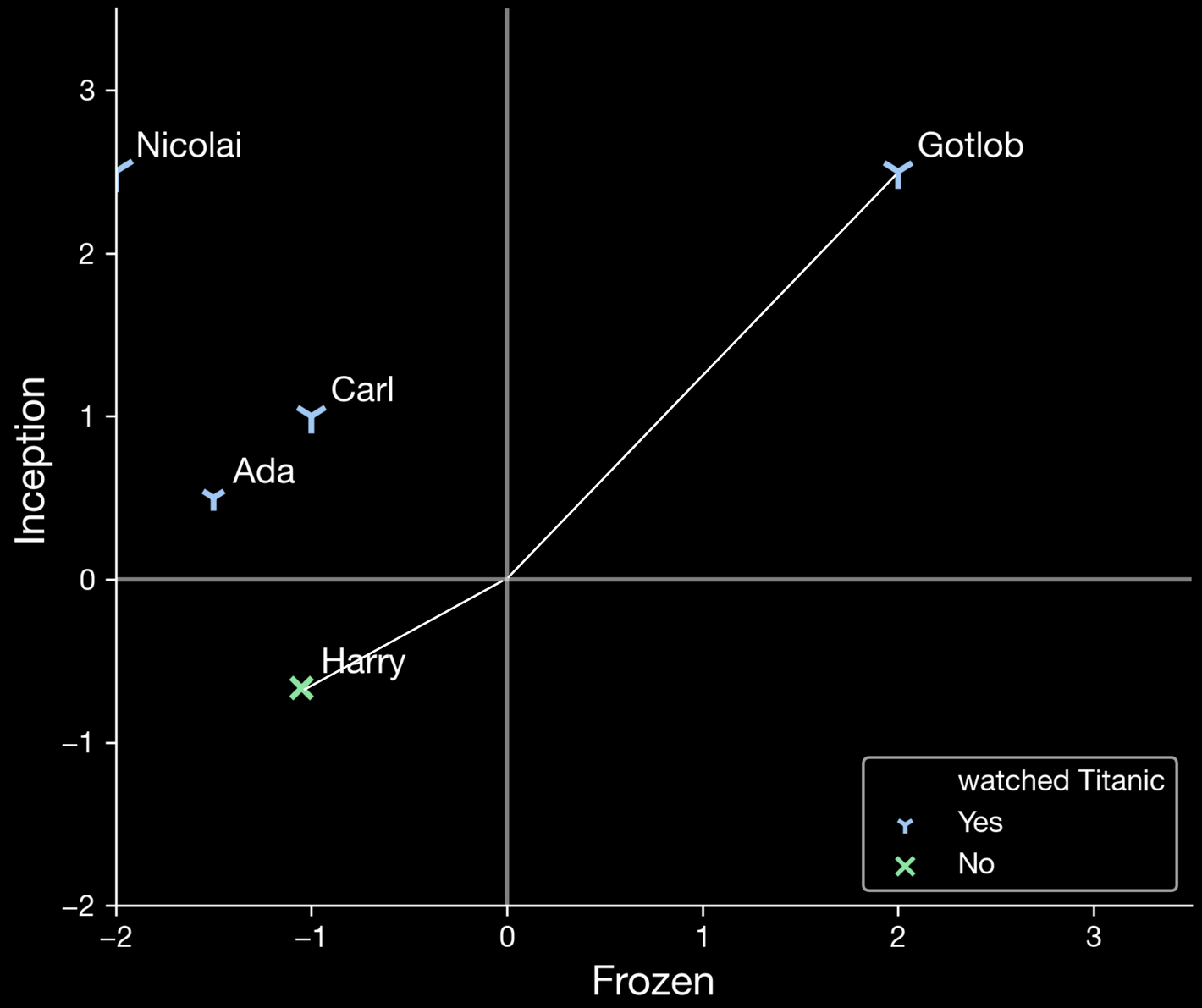


KNN - Movies

Similarity

Sommige mensen ranken altijd positiever dan anderen.

Absolute rating is niet belangrijk, maar afwijking ten opzichte van gemiddeld rating.



KNN - Movies

Similarity

Sommige mensen ranken altijd positiever dan anderen.

Absolute rating is niet belangrijk, maar afwijking ten opzichte van gemiddeld rating.

Gemiddelde rating Harry: μ_{harry}

Relatieve rating van Harry voor Frozen: $r_{harry,frozen} - \mu_{harry}$

Euclidian distance

$$Cos(Harry, Ada) = \frac{\sum_{i \in I_{harry} \cap I_{ada}} r_{harry,i} \cdot r_{ada,i}}{\sum_{i \in I_{harry} \cap I_{ada}} r^2_{harry,i} \cdot \sum_{i \in I_{harry} \cap I_{ada}} r^2_{ada,i}}$$

$$Cos(u, v) = \frac{\sum_{i \in I_u \cap I_v} r_{u,i} \cdot r_{v,i}}{\sum_{i \in I_u \cap I_v} r^2_{u,i} \cdot \sum_{i \in I_u \cap I_v} r^2_{v,i}}$$

Euclidian distance

$$Cos(Harry, Ada) = \frac{\sum_{i \in I_{harry} \cap I_{ada}} (r_{harry,i} - \mu_{harry}) \cdot (r_{ada,i} - \mu_{ada})}{\sum_{i \in I_{harry} \cap I_{ada}} (r_{harry,i} - \mu_{harry})^2 \cdot \sum_{i \in I_{harry} \cap I_{ada}} (r_{ada,i} - \mu_{ada})^2}$$

$$Cos(u, v) = \frac{\sum_{i \in I_u \cap I_v} r_{u,i} \cdot r_{v,i}}{\sum_{i \in I_u \cap I_v} r_{u,i}^2 \cdot \sum_{i \in I_u \cap I_v} r_{v,i}^2}$$

Euclidian distance

$$Cos(Harry, Ada) = \frac{\sum_{i \in I_{harry} \cap I_{ada}} (r_{harry,i} - \mu_{harry}) \cdot (r_{ada,i} - \mu_{ada})}{\sum_{i \in I_{harry} \cap I_{ada}} (r_{harry,i} - \mu_{harry})^2 \cdot \sum_{i \in I_{harry} \cap I_{ada}} (r_{ada,i} - \mu_{ada})^2}$$

$$Cos(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u) \cdot (r_{v,i} - \mu_v)}{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u)^2 \cdot \sum_{i \in I_u \cap I_v} (r_{v,i} - \mu_v)^2}$$

$$Cos(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u) \cdot (r_{v,i} - \mu_v)}{\sum_{i \in I_u \cap I_v} (r_{u,i} - \mu_u)^2 \cdot \sum_{i \in I_u \cap I_v} (r_{v,i} - \mu_v)^2}$$