Machine Learning
Algorithms

Common ML Algorithms

Linear Regression

Logistic Regression

Support Vector Machine

Decision Tree

K-Nearest Neighbor

Neural Network

$$\hat{f}(X)$$

The Prediction

output input

$$\hat{y} = \hat{f}(X)$$

Regression

continuous

$$\hat{y} = 66.5$$

Classification

probability

$$\hat{y} = .85$$

Classification

prediction

$$\hat{y} = 1$$

Classification

prediction

$$\hat{y} = 0$$

Linear Regression

equation of a line

$$y = mx + b$$

equation of a line

$$y = mx + b$$

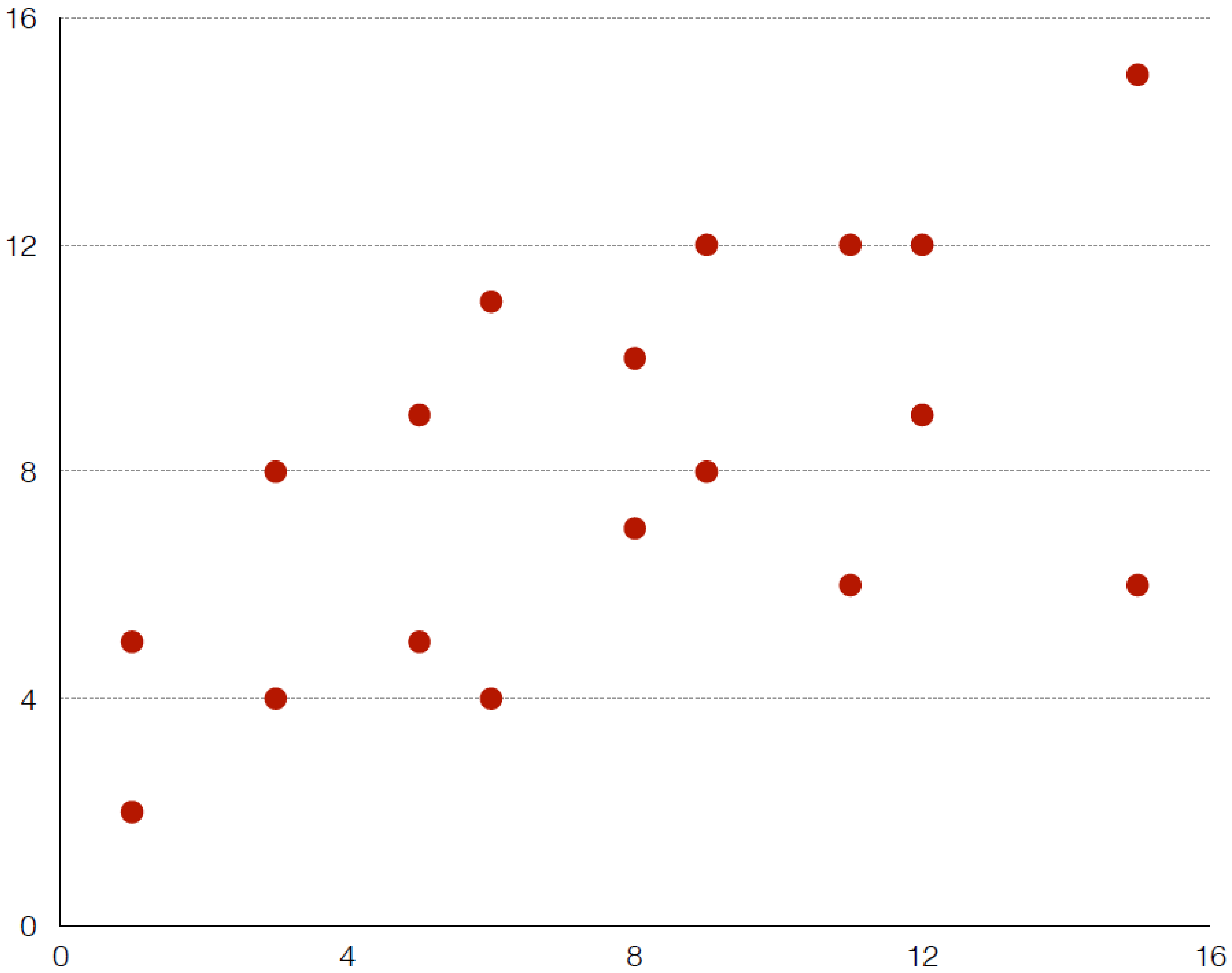
linear regression

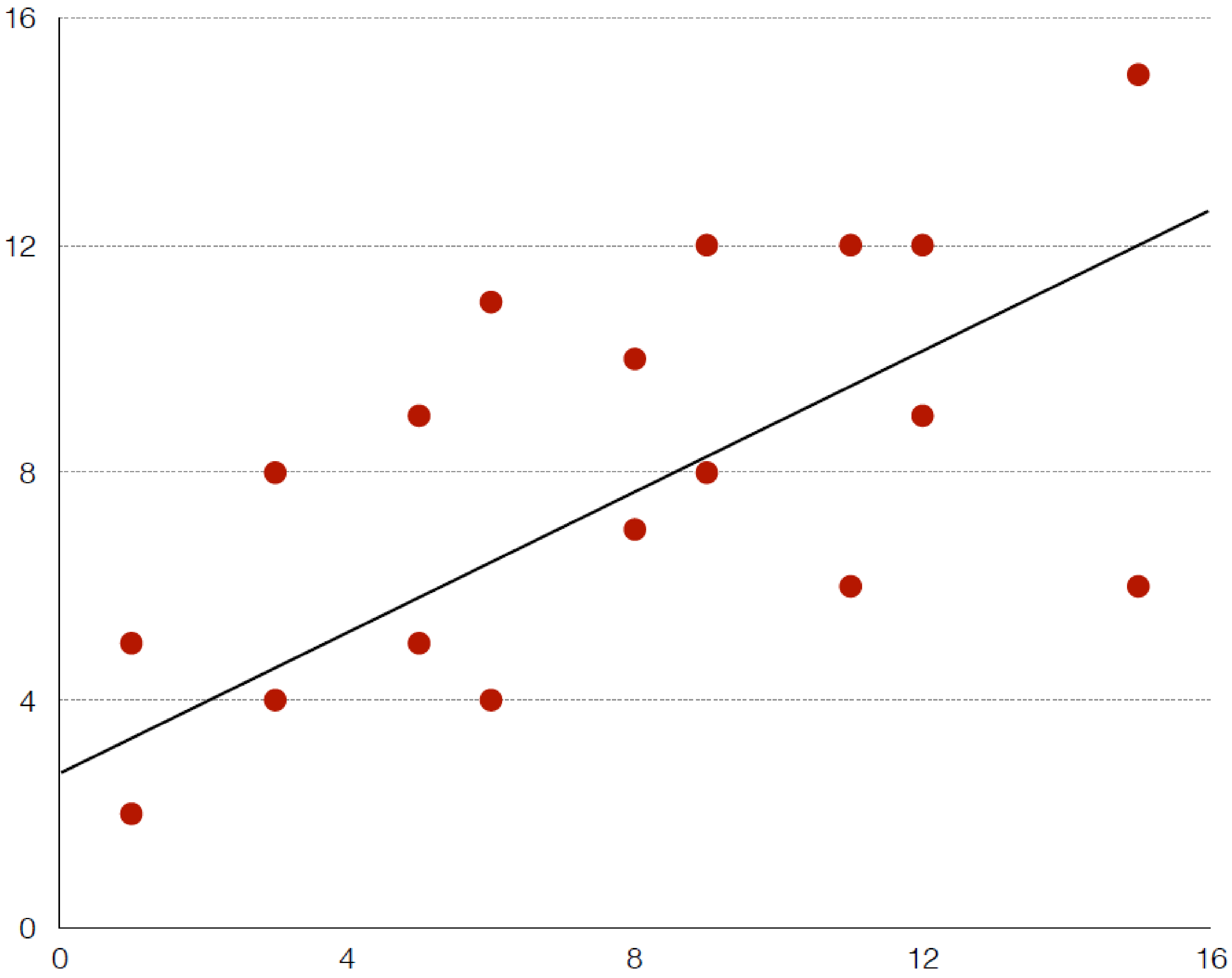
$$y = \beta_0 + \beta_1 x$$

Simple Linear Regression

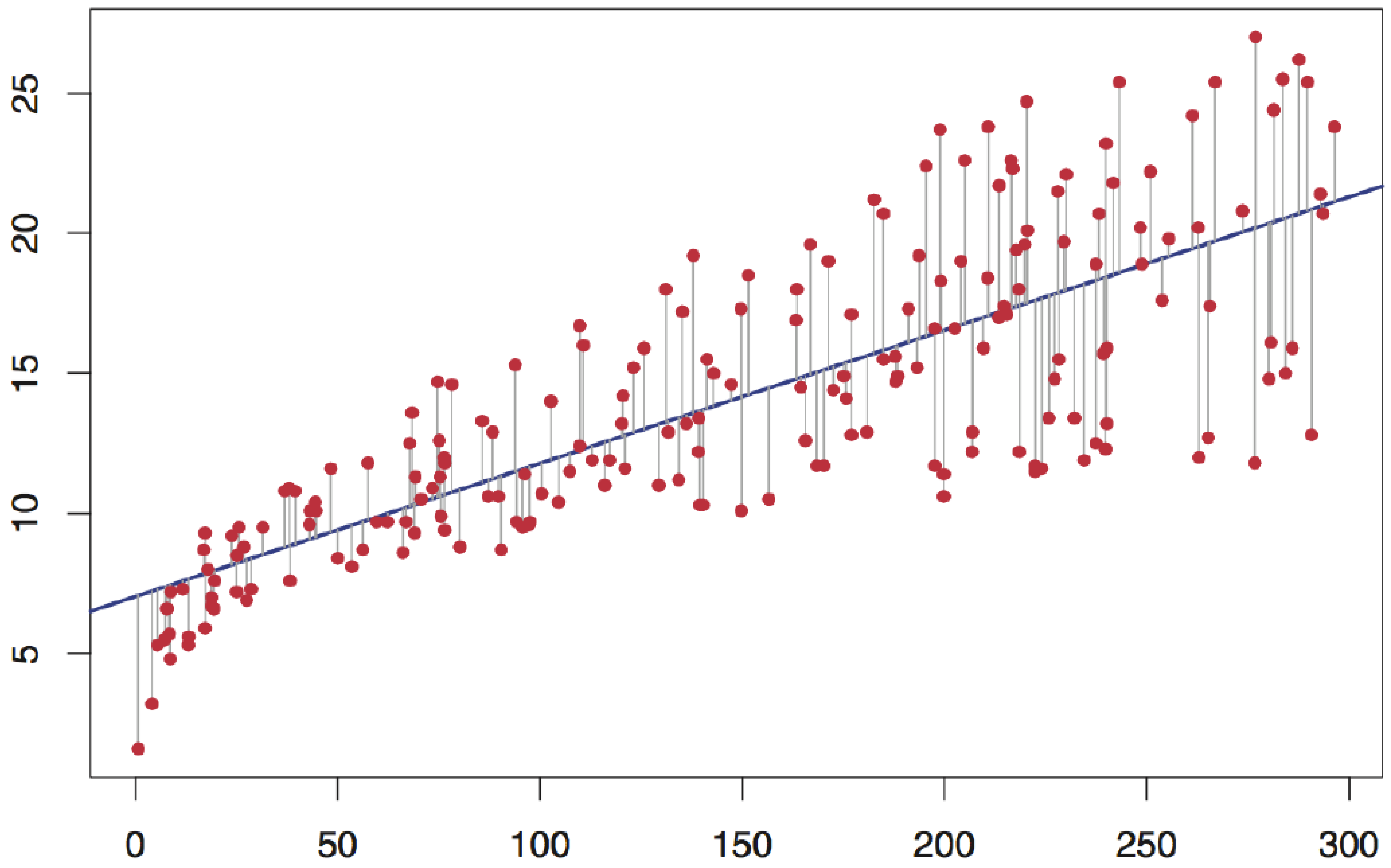
Input	learned coefficients (weights)	output
x_1	β_0, β_1	y

$$y = \beta_0 + \beta_1 x$$


16
12
8
4
0
0
4
8
12
16



16
12
8
4
0
0
4
8
12
16



5
10
15
20
25
0
50
100
150
200
250
300

Weekly Hours Spent Studying

Gender	Number of Classes	Social Accounts	Hours Studying
Male	2	1	7.5
Male	4	3	12.25
Female	4	3	12.75
Female	3	4	7.75
Female	4	2	14
Male	2	3	5.75
Female	5	1	18.25

Weekly Hours Spent Studying

Gender	Number of Classes	Social Accounts	Hours Studying
0	2	1	7.5
0	4	3	12.25
1	4	3	12.75
1	3	4	7.75
1	4	2	14
0	2	3	5.75
0	5	1	18.25

Weekly Hours Spent Studying

Features
(X)

Gender	Number of Classes	Social Accounts	Hours Studying
0	2	1	7.5
0	4	3	12.25
1	4	3	12.75
1	3	4	7.75
1	4	2	14
0	2	3	5.75
0	5	1	18.25

Weekly Hours Spent Studying

Gender	Number of Classes	Social Accounts	Hours Studying
0	2	1	7.5
0	4	3	12.25
1	4	3	12.75
1	3	4	7.75
1	4	2	14
0	2	3	5.75
0	5	1	18.25

Target
(y)

Weekly Hours Spent Studying

	Gender	Number of Classes	Social Accounts	Hours Studying
	0	2	1	7.5
	0	4	3	12.25
Sample →	1	4	3	12.75
	1	3	4	7.75
	1	4	2	14
	0	2	3	5.75
	0	5	1	18.25

Weekly Hours Spent Studying

Gender	Number of Classes	Social Accounts	Hours Studying
0	2	1	7.5
0	4	3	12.25
1	4	3	12.75
1	3	4	7.75
1	4	2	14
0	2	3	5.75
0	5	1	18.25

Weekly Hours Spent Studying

0	2	1	7.5
0	4	3	12.25
1	4	3	12.75
1	3	4	7.75
1	4	2	14
0	2	3	5.75
0	5	1	18.25

Weekly Hours Spent Studying

$$\mathbf{X} = \begin{matrix} 0 & 2 & 1 \\ 0 & 4 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 2 \\ 0 & 2 & 3 \\ 0 & 5 & 1 \end{matrix} \qquad \mathbf{y} = \begin{matrix} 7.5 \\ 12.25 \\ 12.75 \\ 7.75 \\ 14 \\ 5.75 \\ 18.25 \end{matrix}$$

Multiple Linear Regression

$$\textbf{Hrs. Studying} = 1.63 + 3.51x_1 + .25x_2 - 1.08x_3$$

x_1 = number of classes

x_2 = gender (m=0, female=1)

x_3 = number of social accounts

Multiple Linear Regression

$$\textbf{Apps Sold} = 46.55 + 35.03x_1 + 7.11x_2 + 52.48x_3$$

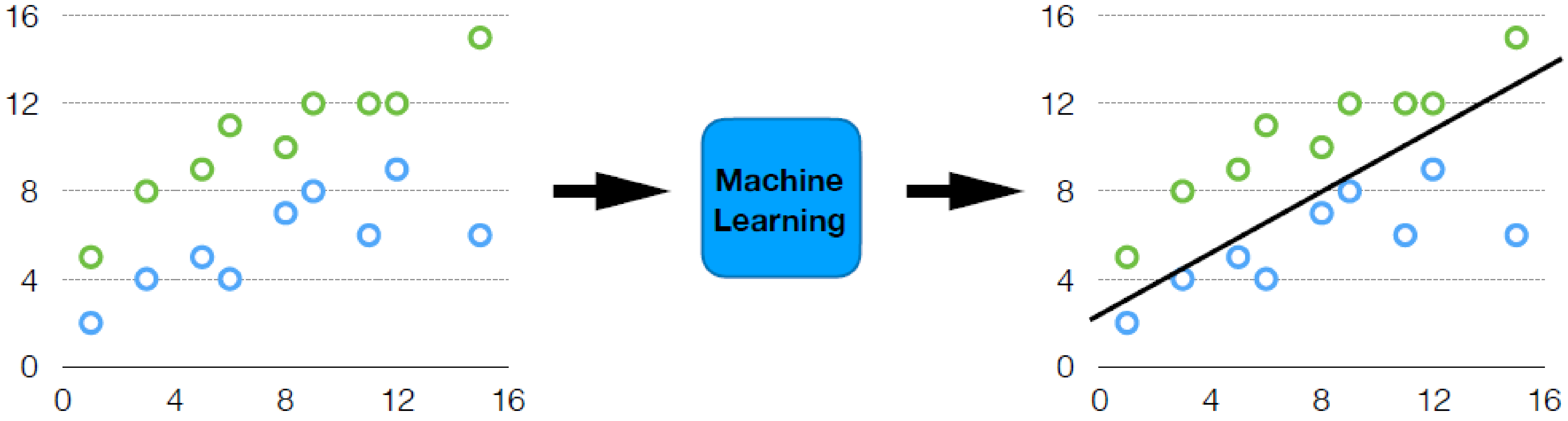
x_1 = per $100 of advertising

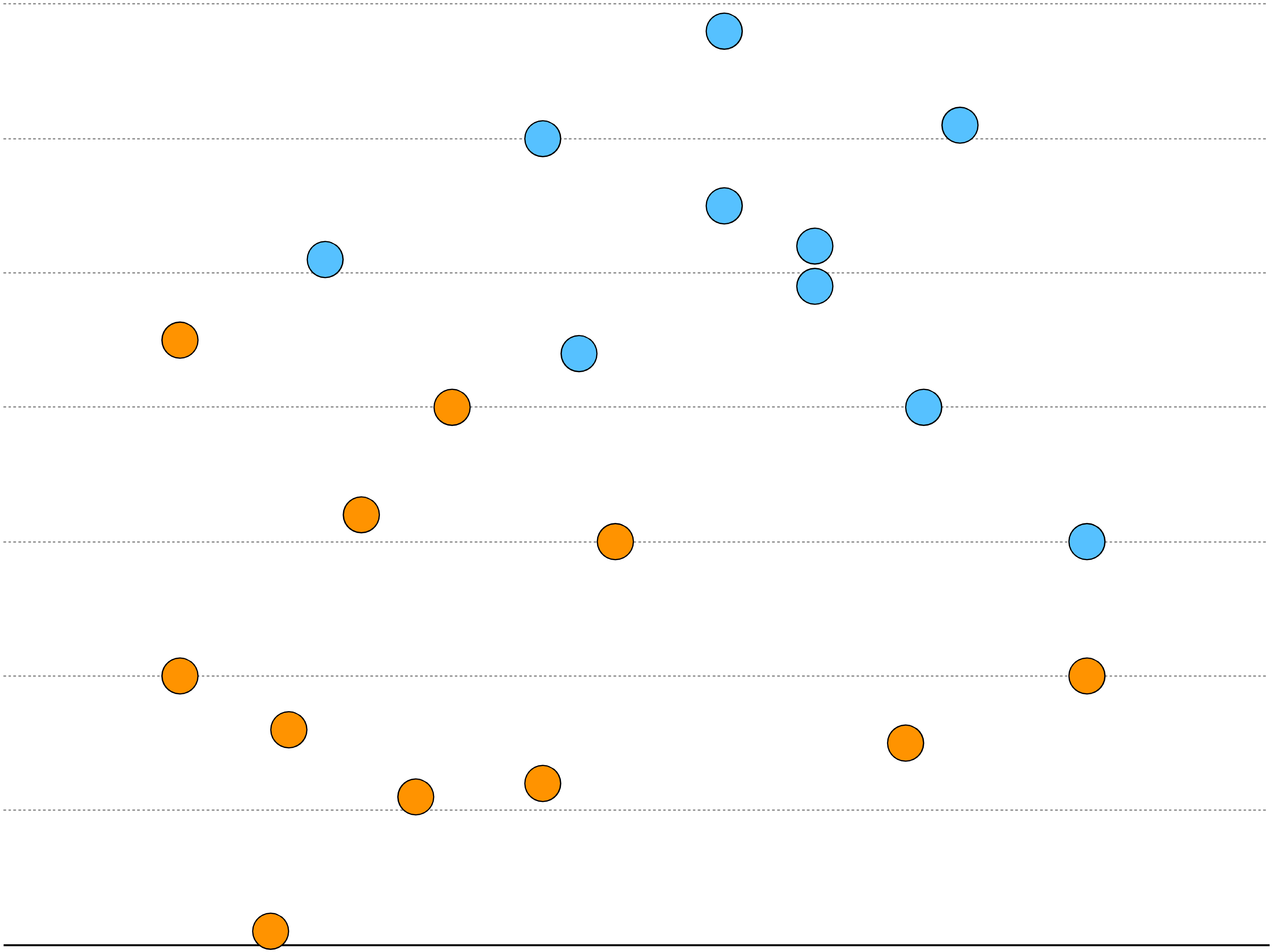
x_2 = public talks

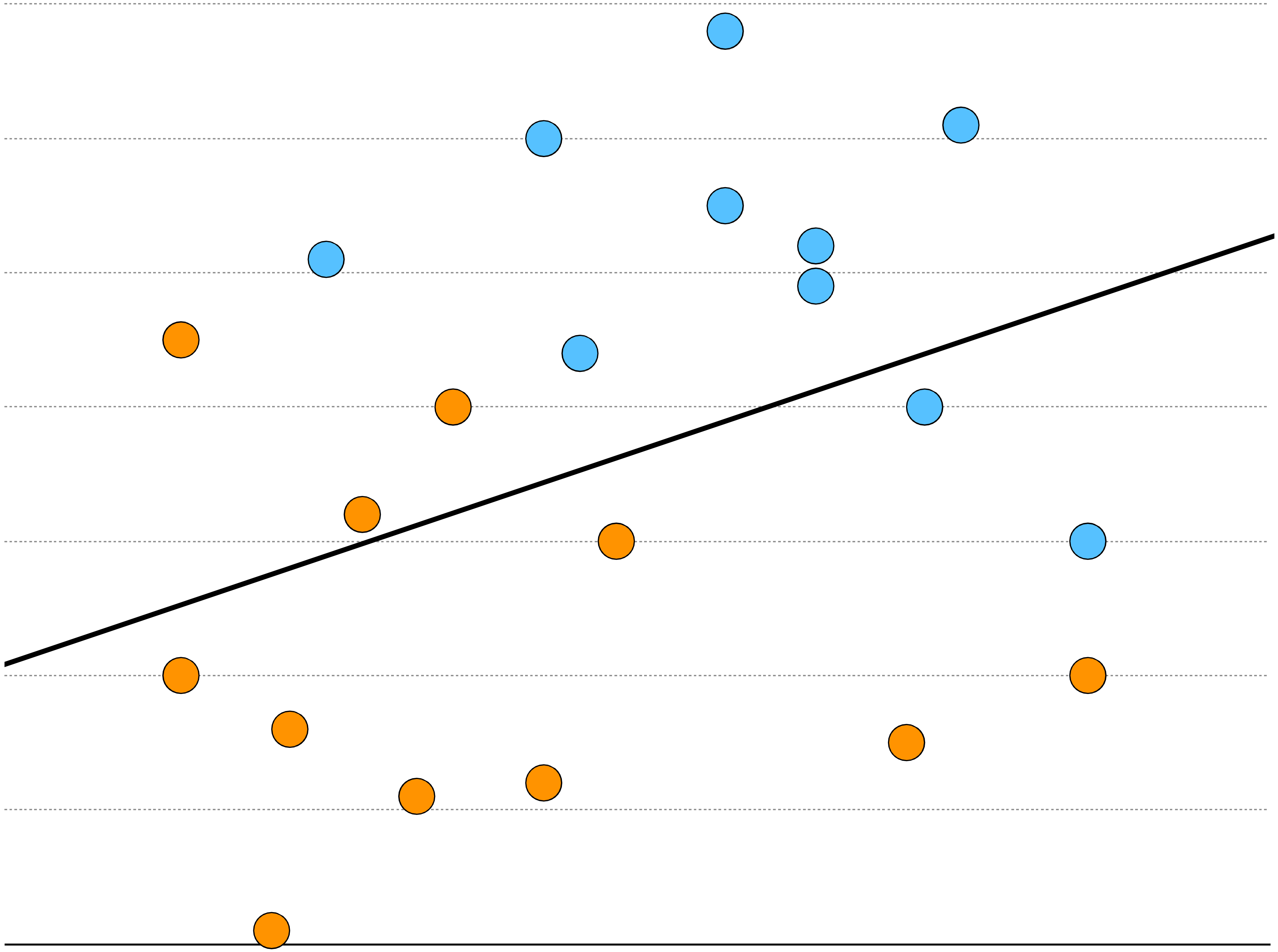
x_3 = targeted podcasts

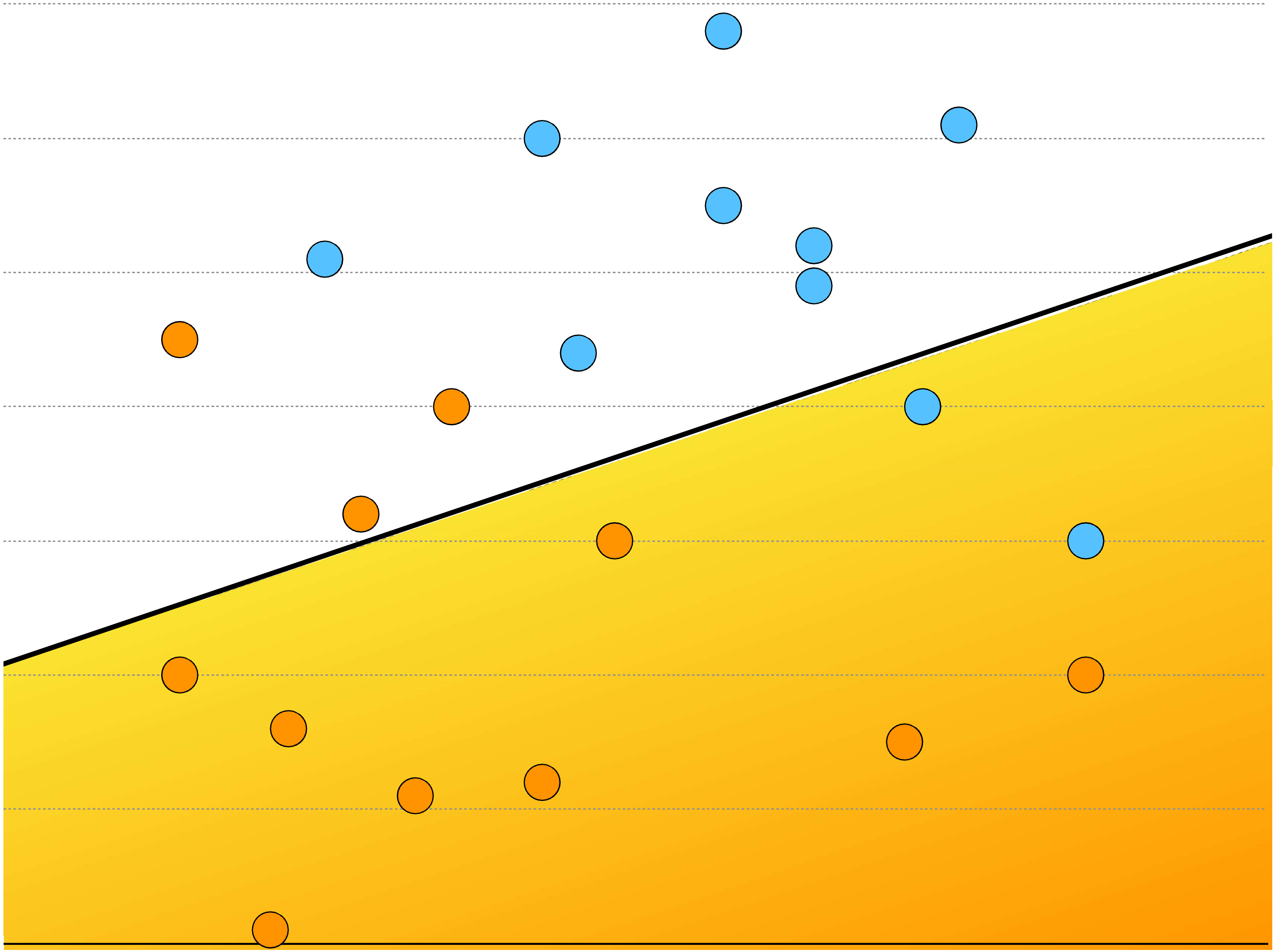
Classification

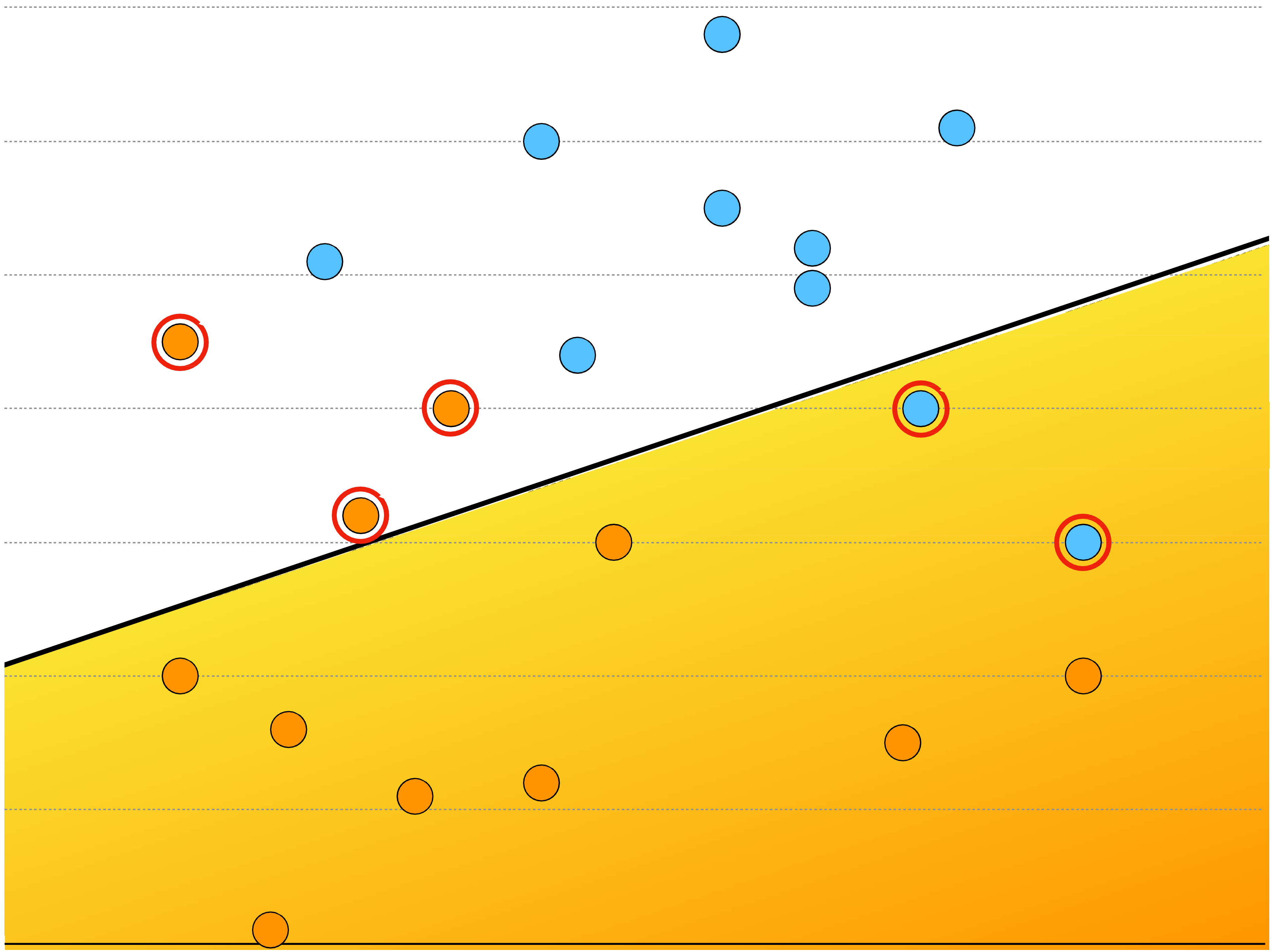
What is a classifier?

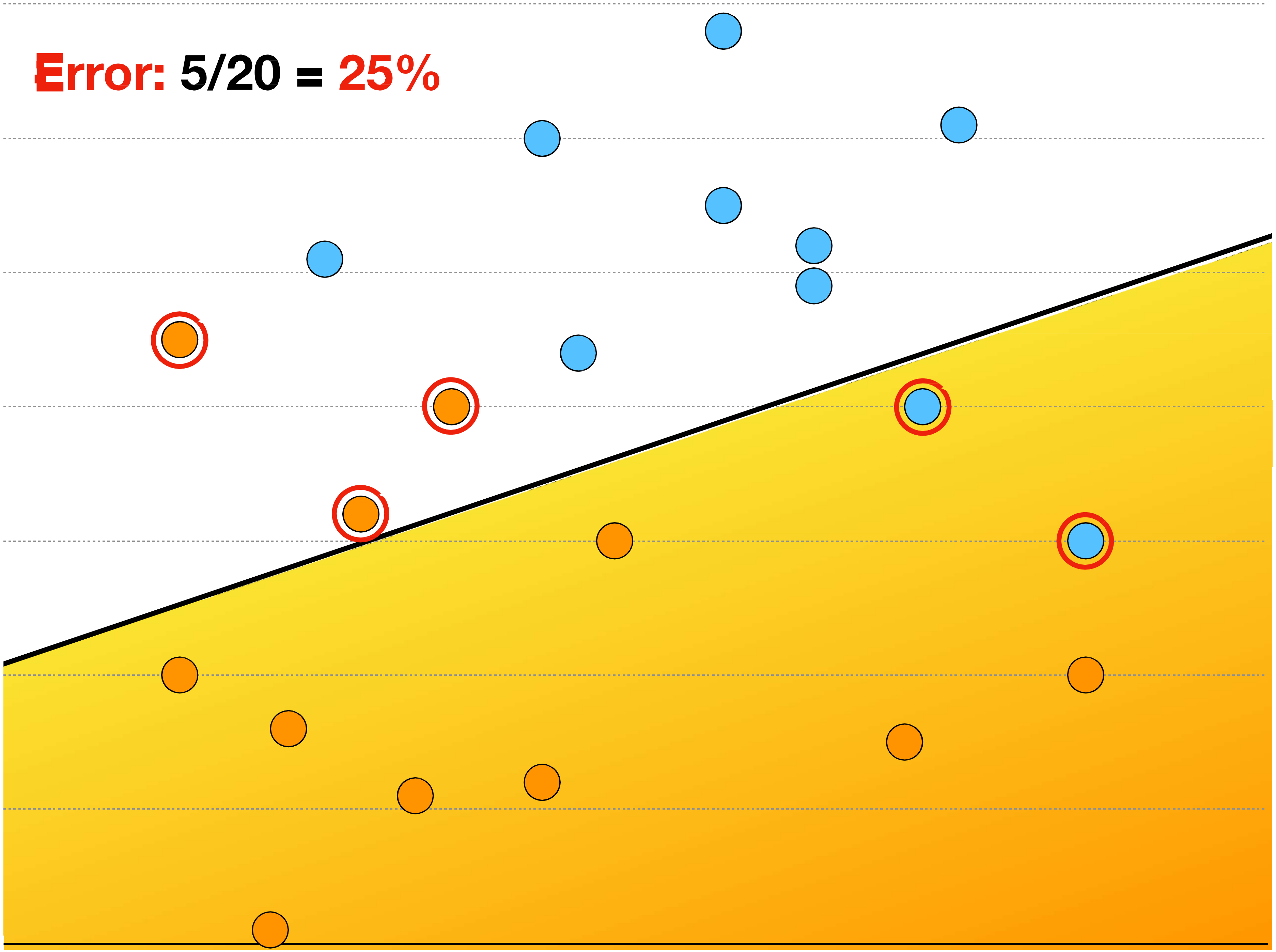


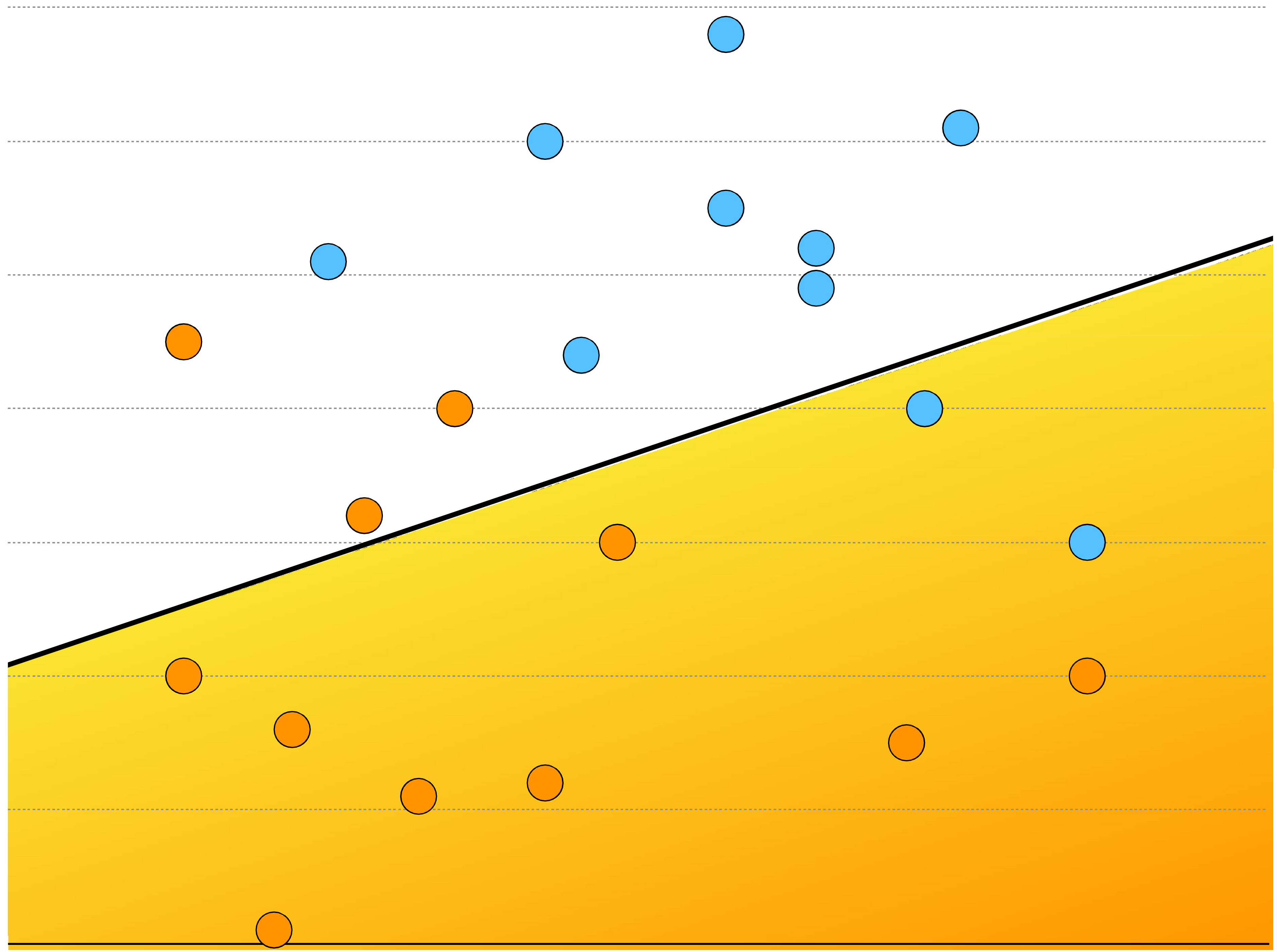


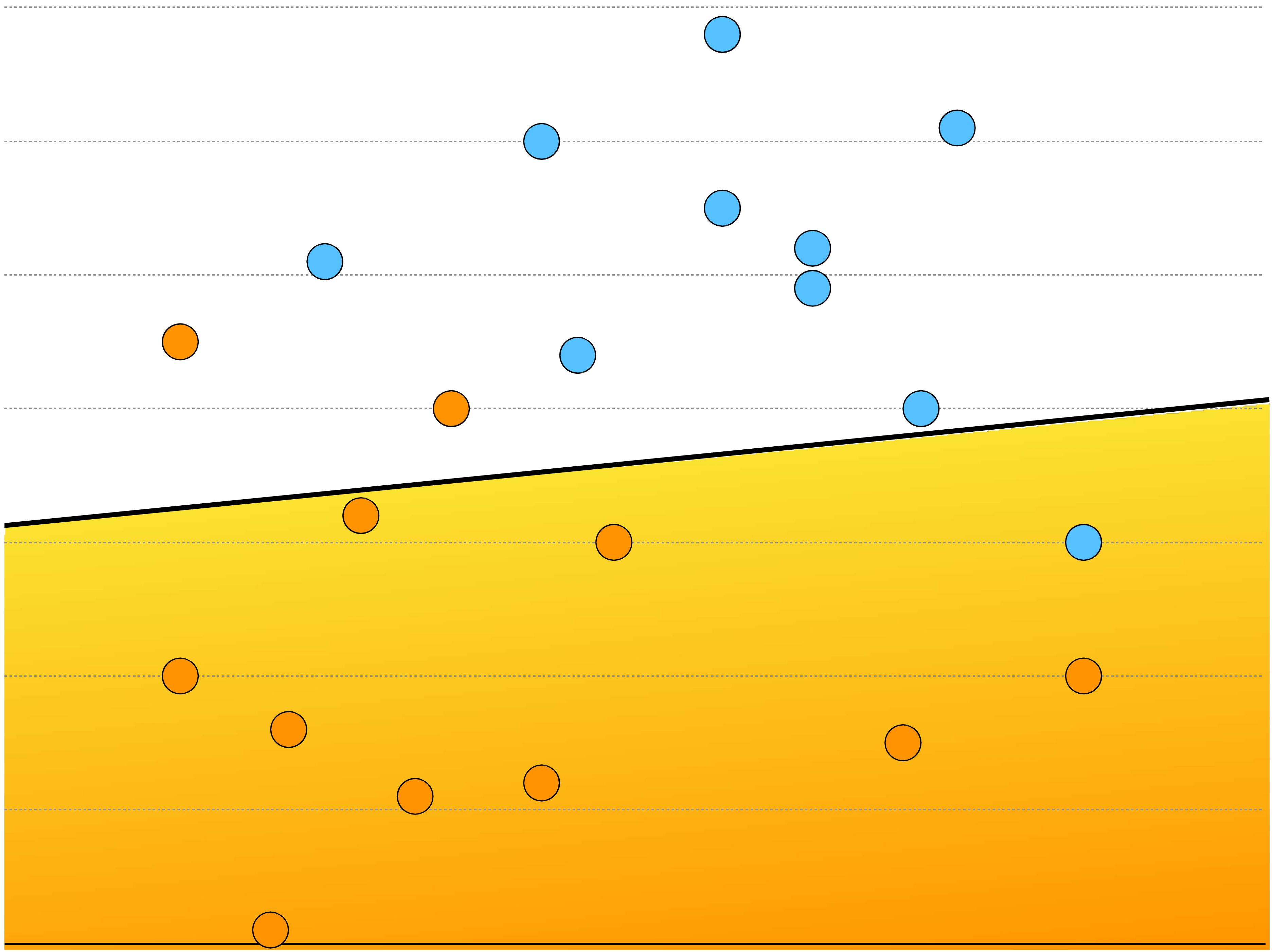


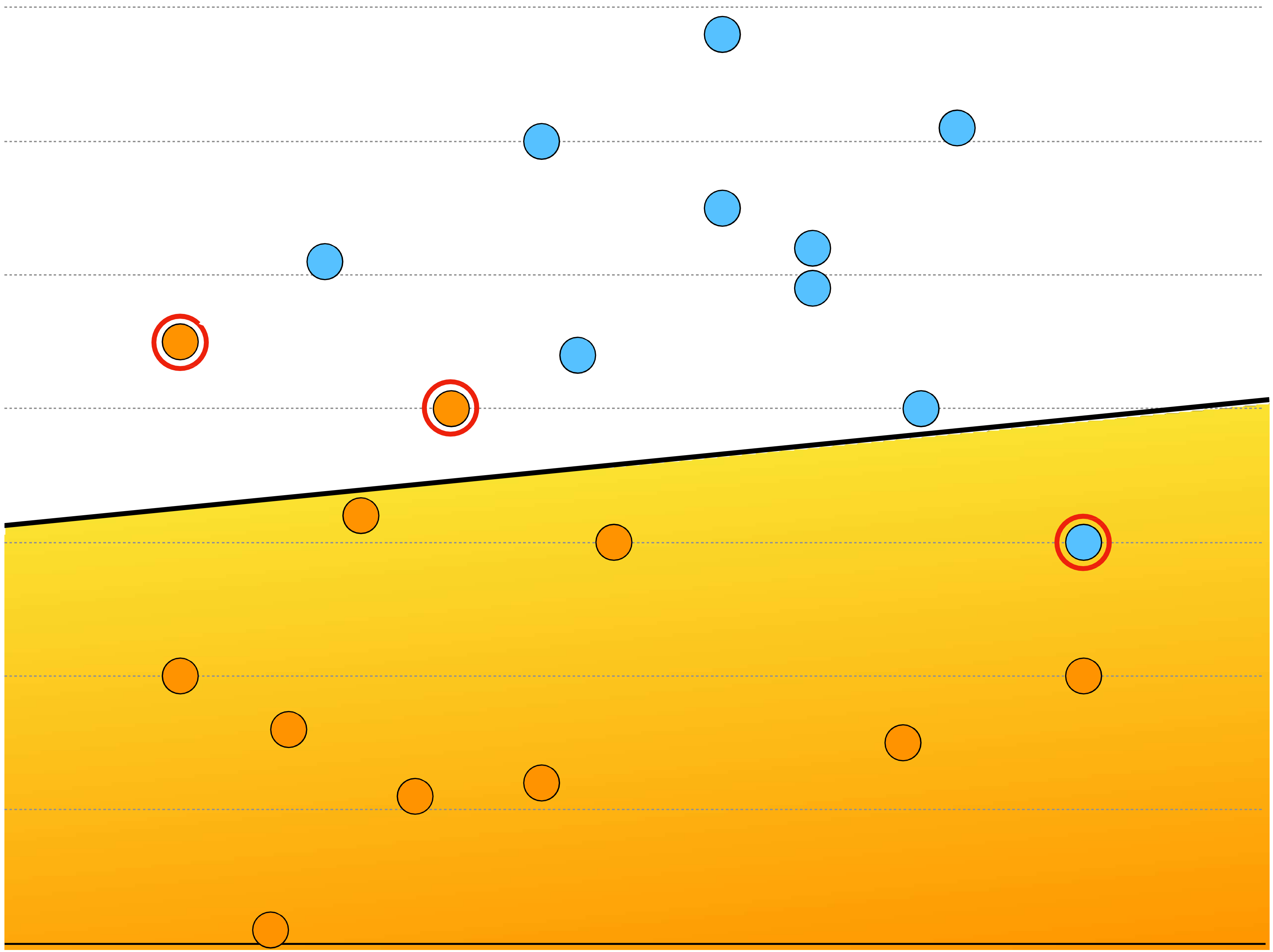


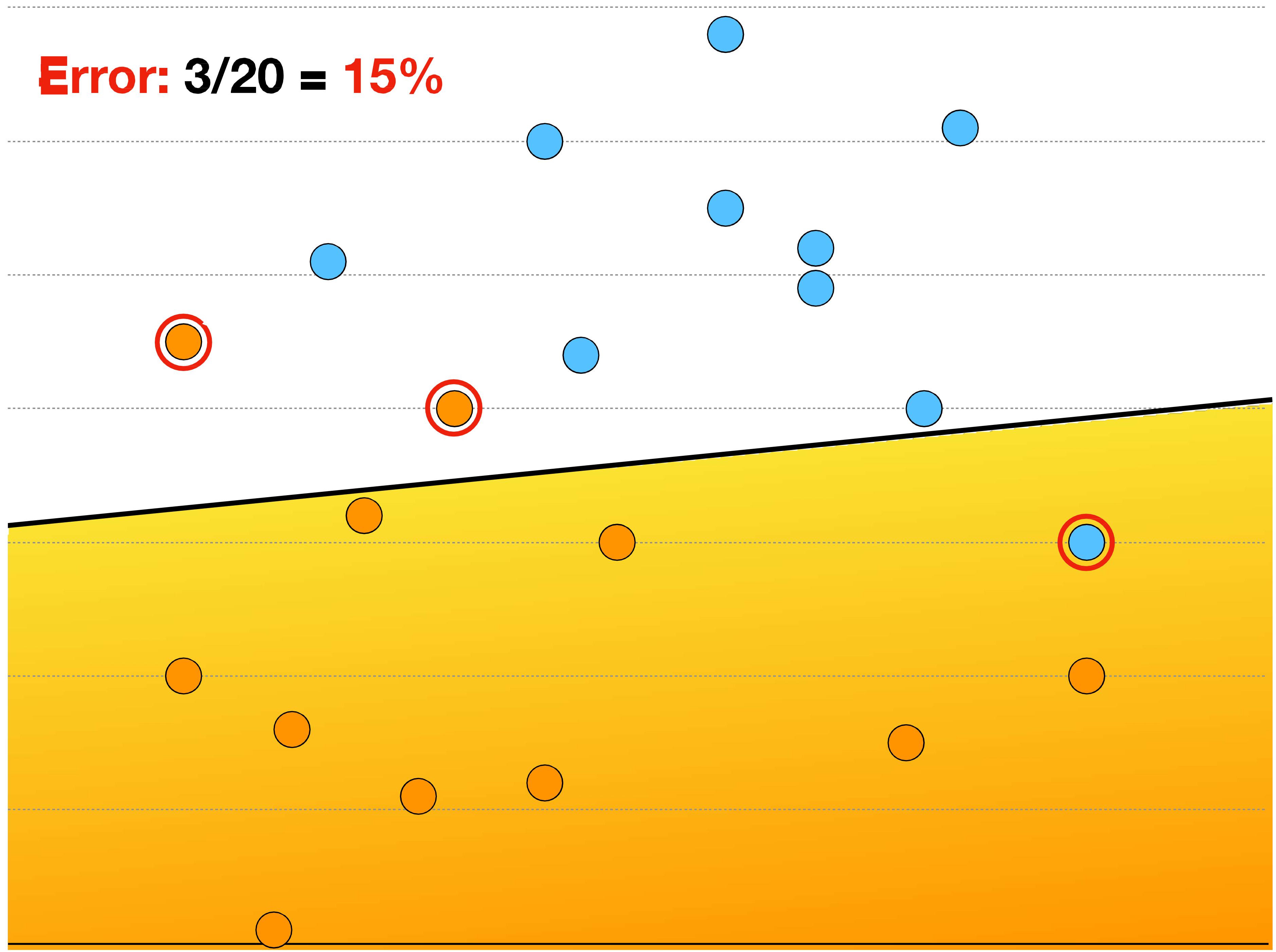


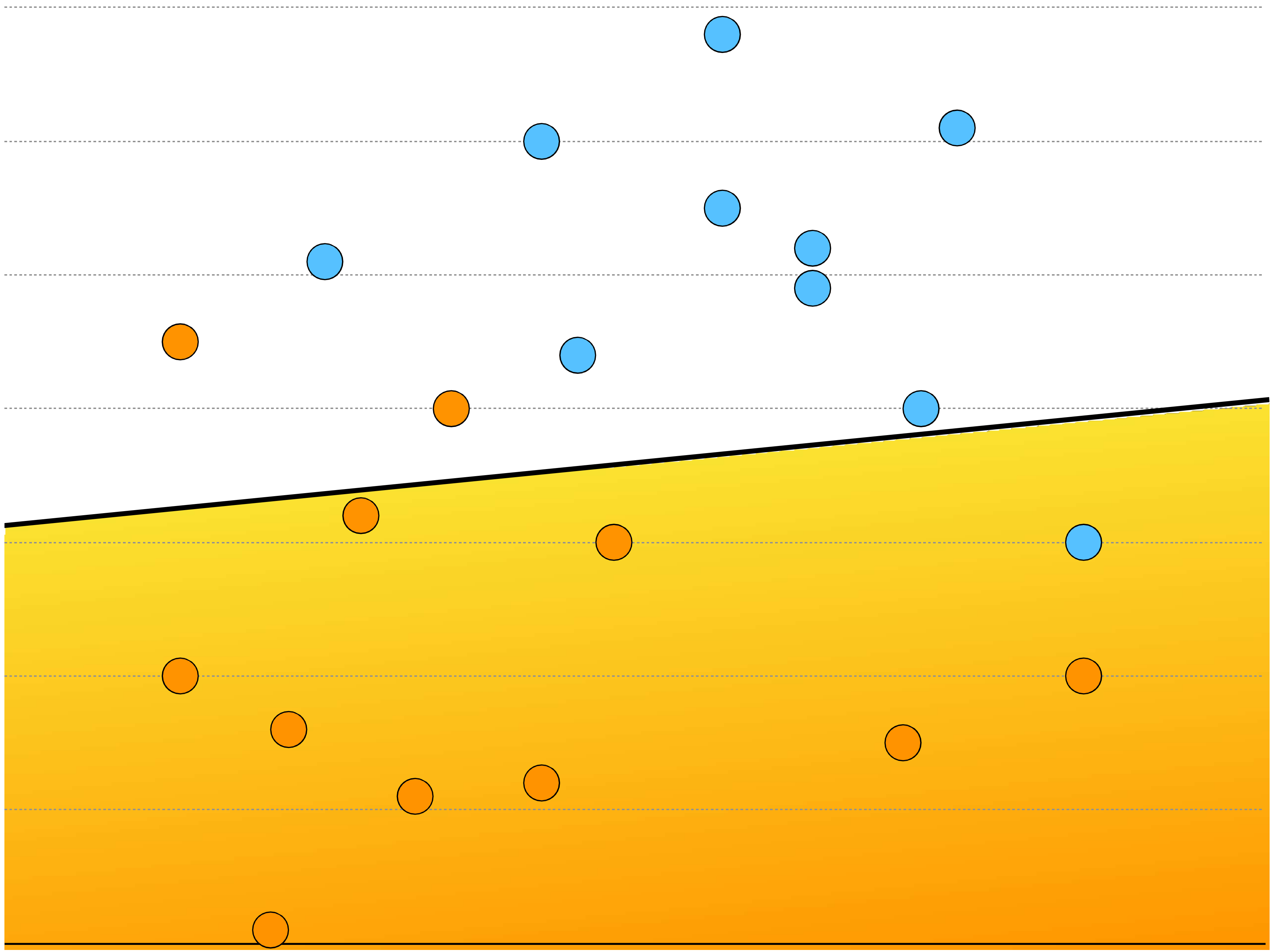

Error: 5/20 = 25%

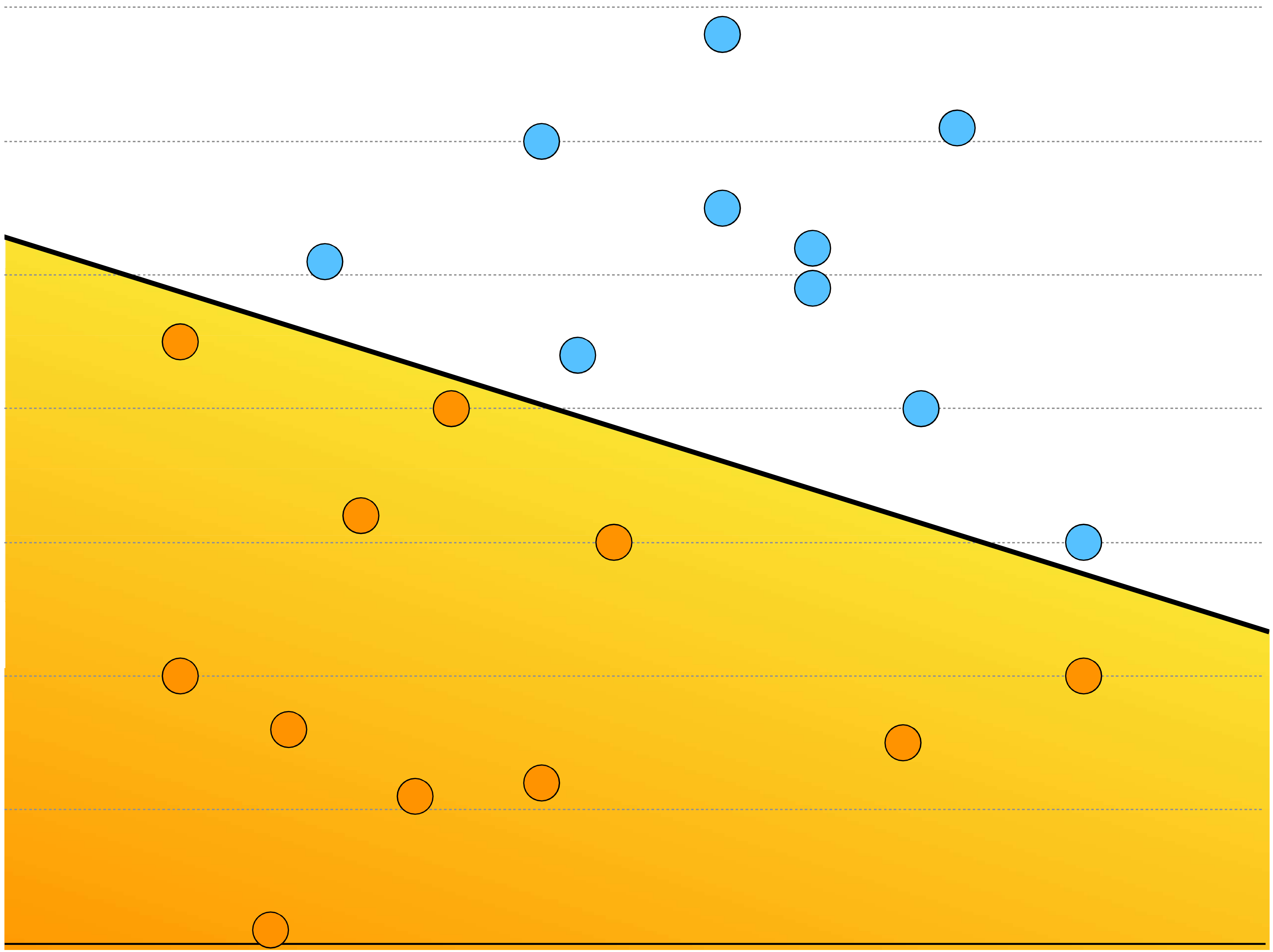


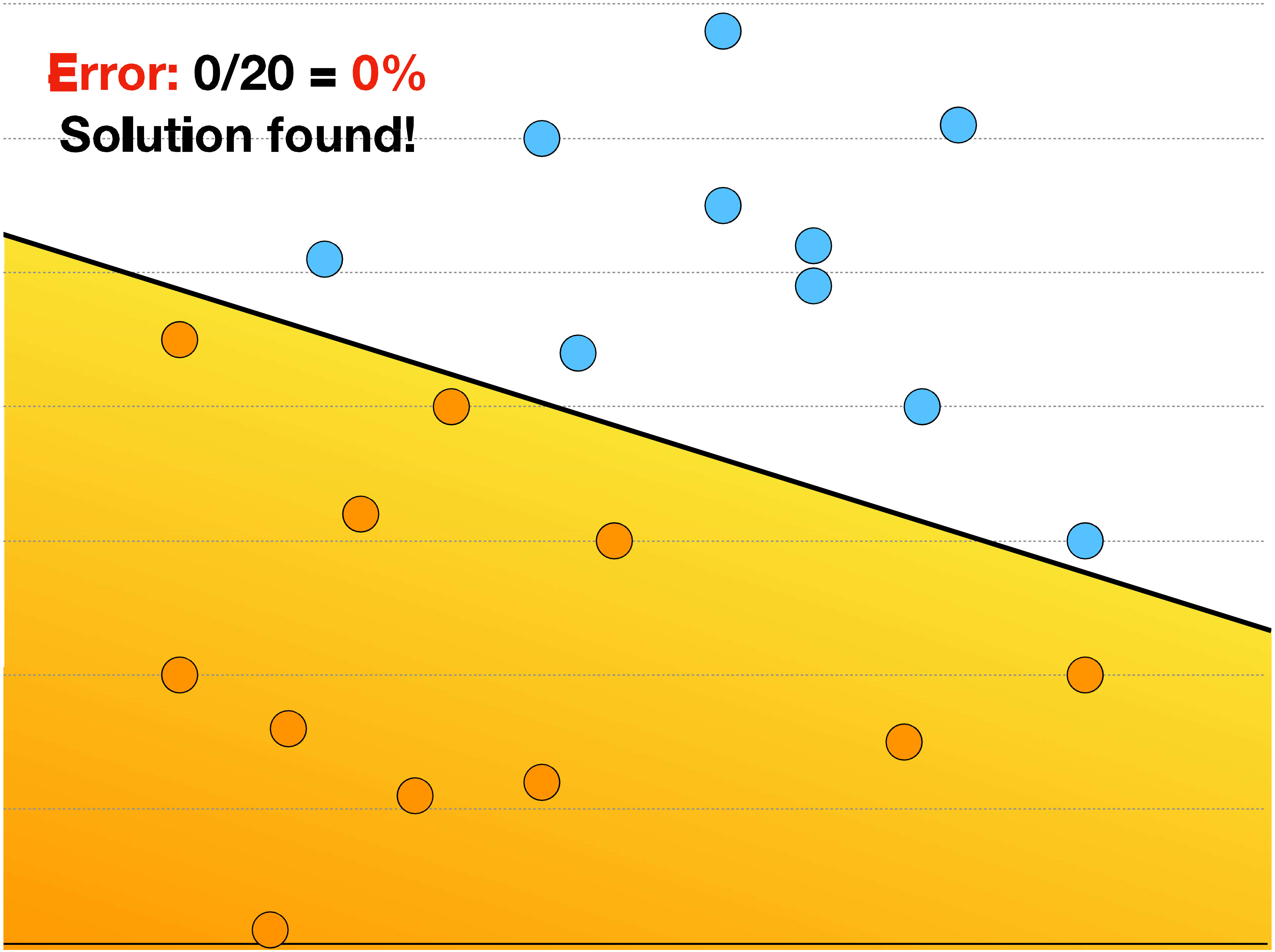





Error: 3/20 = 15%






Error: 0/20 = 0%
Solution found!

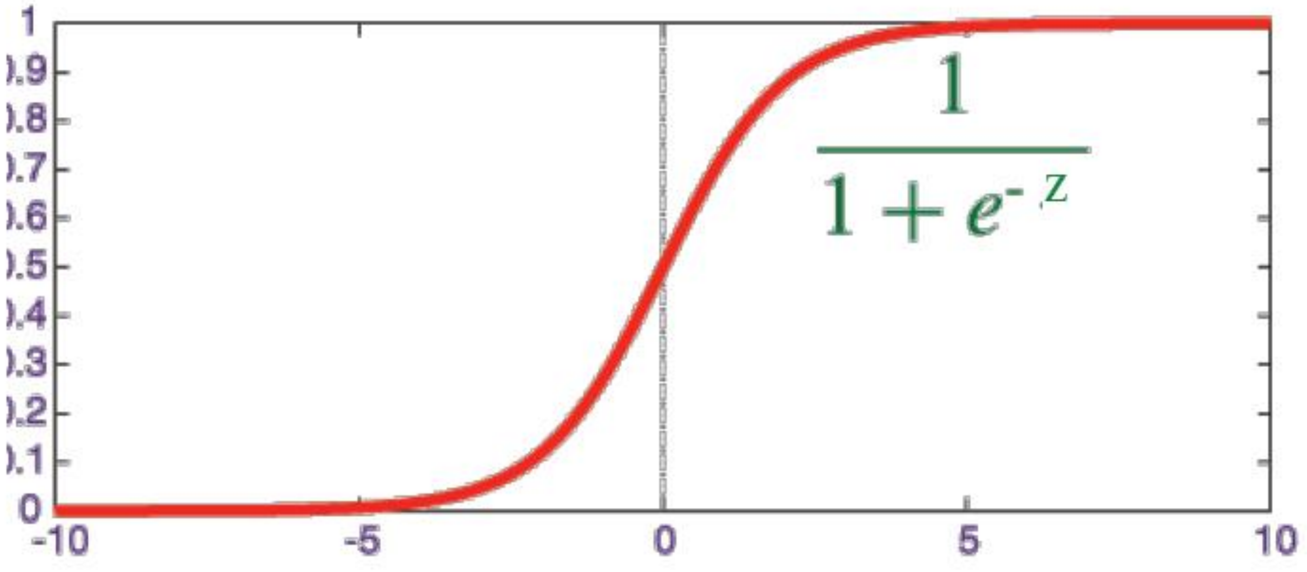
Logistic Regression

Logistic Regression

intermediate step

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Logistic (Sigmoid) Function

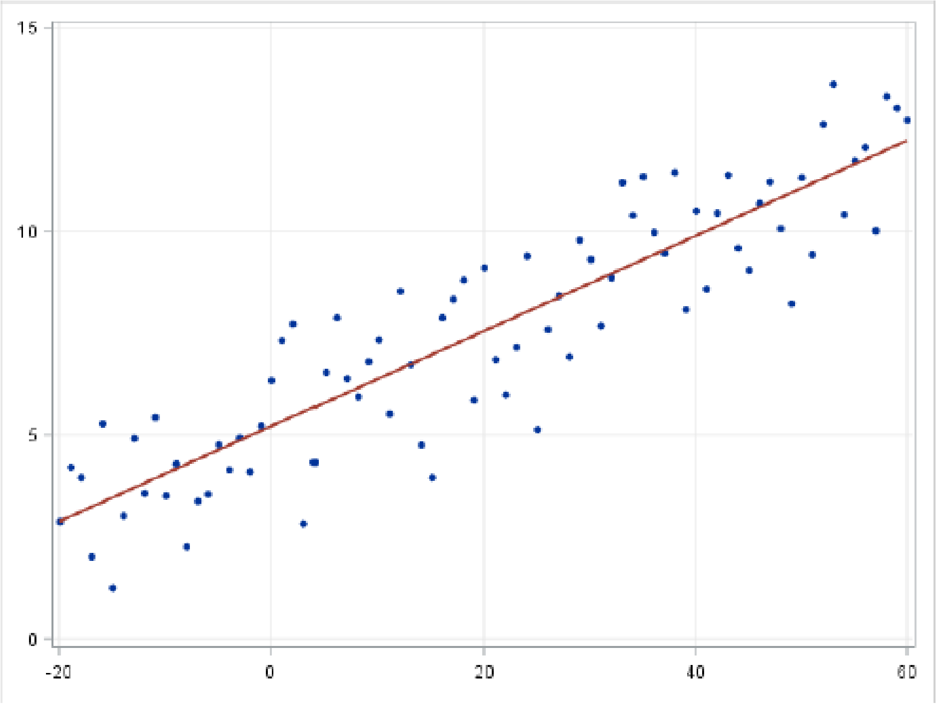


Z

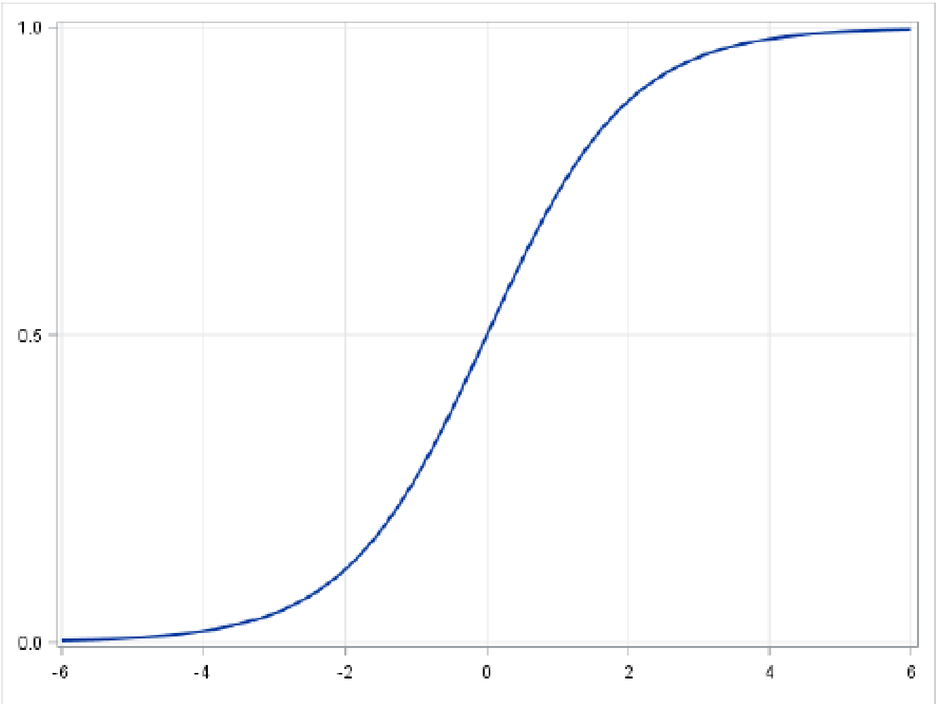
Logistic Regression

$$\hat{f}(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

Linear Regression



Logistic Regression



positive class = 1

negative class = 0

predict 1 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 >= 0$

predict 0 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 < 0$

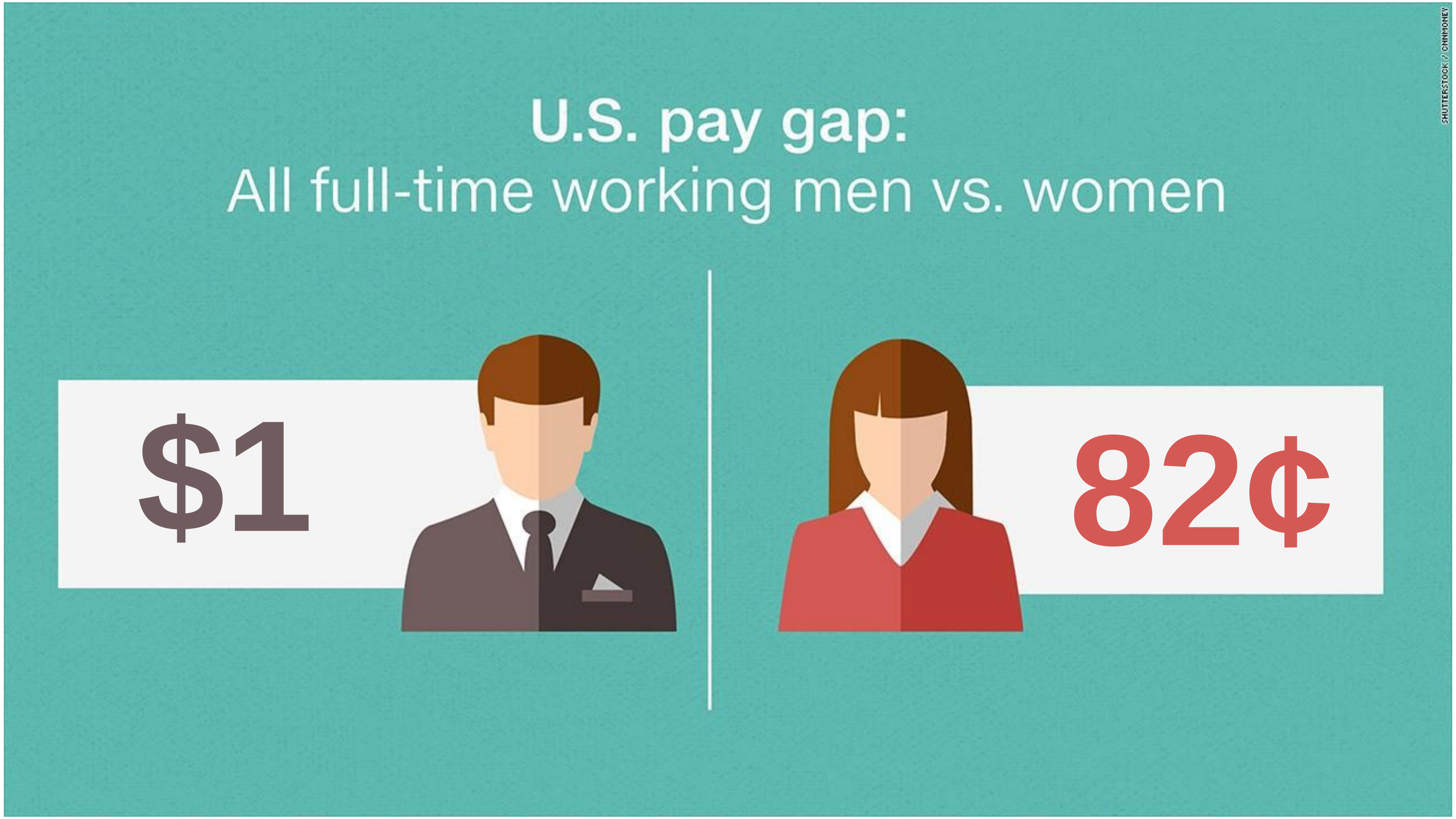
positive class = 1

negative class = 0

if probability >= 0.5: predict 1

if probability < 0.5: predict 0

U.S. Pay Gap


U.S. pay gap:
All full-time working men vs. women
$1
82¢
SHUTTERSTOCK / CNNMONEY

U.S. pay gap:

All full-time working men and women

The highest paid occupations for women

Chief executives

Pharmacists

Lawyers

Computer and information systems managers

Physicians and surgeons

Nurse practitioners

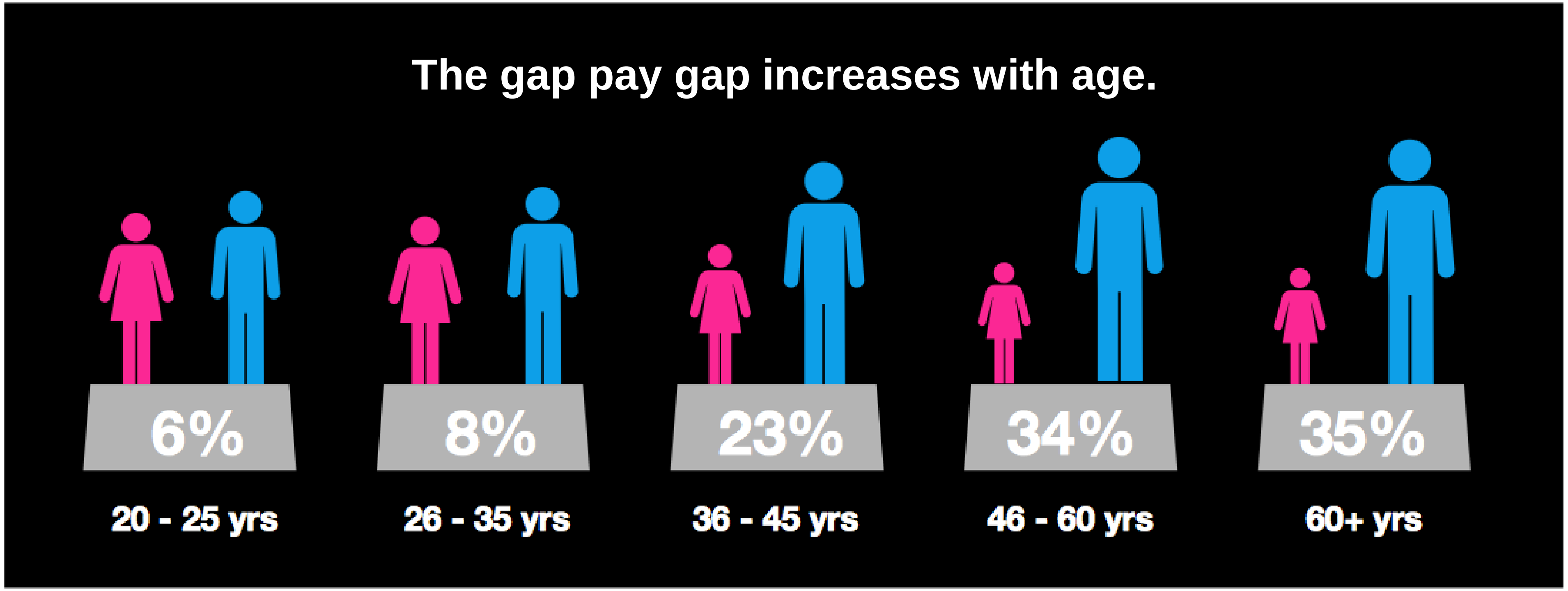
Engineers

Software developers

Management analysts

Operations research analysts

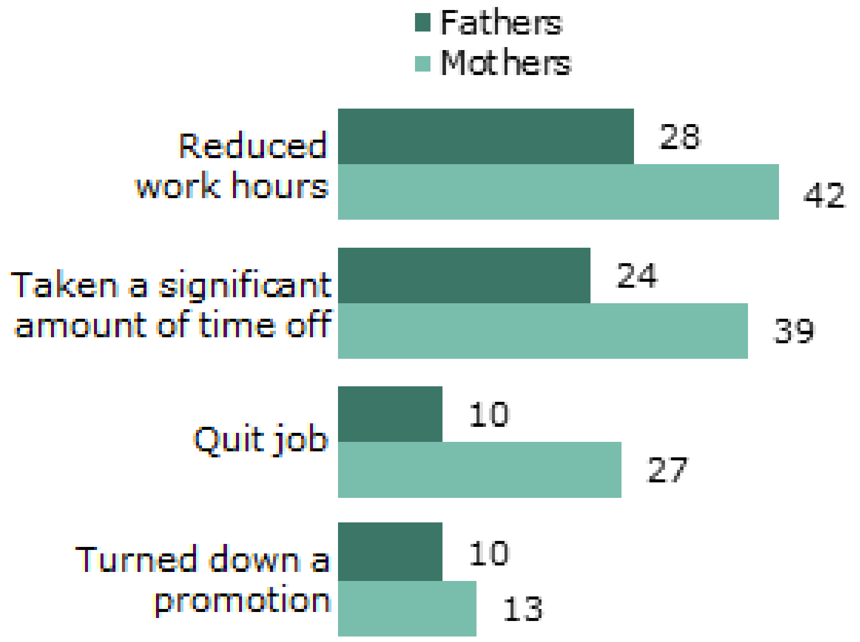
Median Weekly Earnings, by Age and Gender, 2016



The gap pay gap increases with age.
6%
20 - 25 yrs
8%
26 - 35 yrs
23%
36 - 45 yrs
34%
46 - 60 yrs
35%
60+ yrs

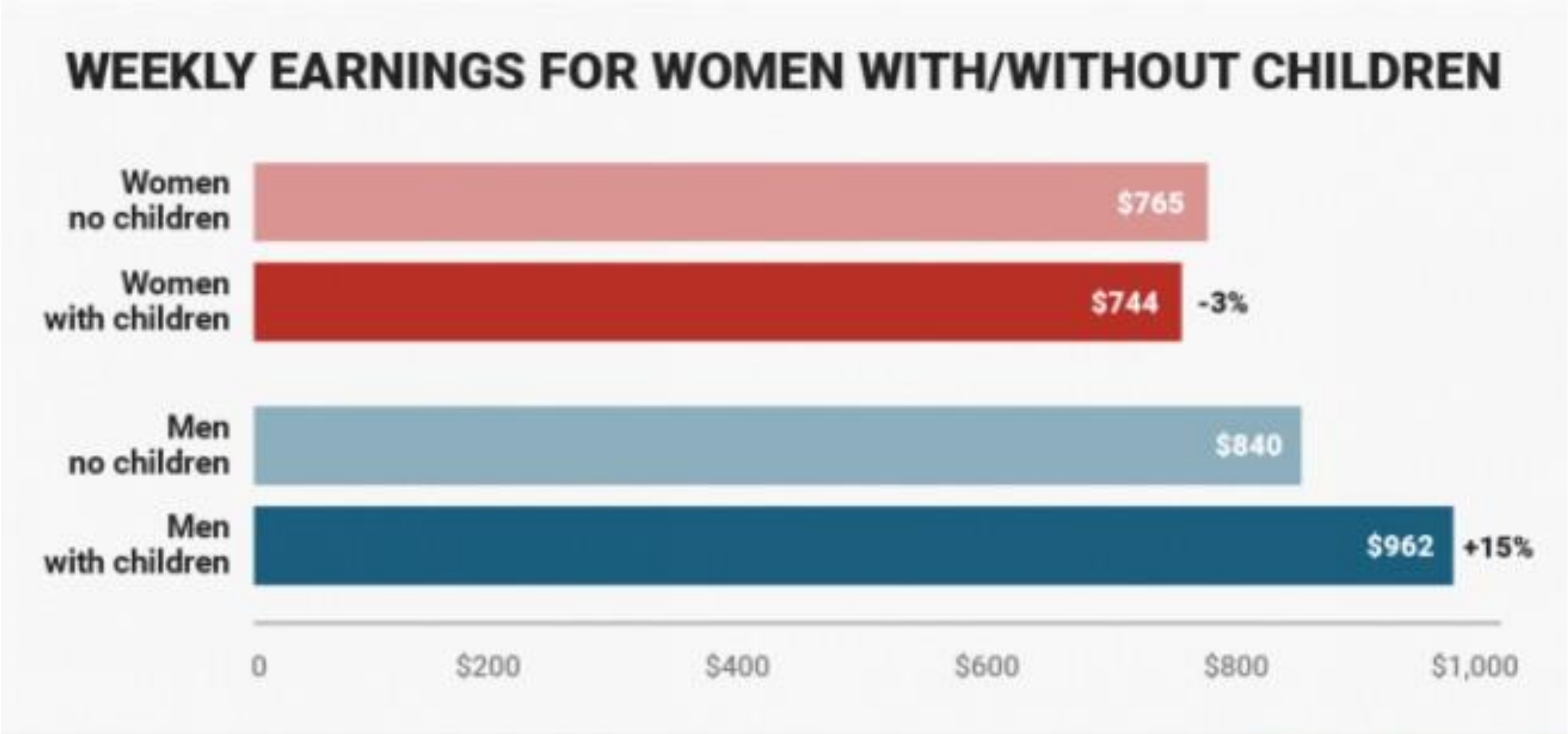
Mothers, More than Fathers, Experience Career Interruptions

% saying they have ... in order to care for a child or family member



Notes: Based on those who have ever worked, "Fathers" and "mothers" include those with children of any age, including adult children (n=1,254).

PEW RESEARCH CENTER Q56a-d


WEEKLY EARNINGS FOR WOMEN WITH/WITHOUT CHILDREN
Women
no children
$765
Women
with children
$744
-3%
Men
no children
$840
Men
with children
$962
+15%
0
$200
$400
$600
$800
$1,000

Data

Race 1	Race 2	Race 3	Race 4	Top Paying	Age	Children	Pay Gap
1	0	0	0	0	49	1	1
0	0	0	0	1	43	0	0
0	0	1	0	0	37	0	0
0	1	0	0	0	28	1	1
0	1	0	0	0	25	1	0
0	0	0	1	0	23	0	0
0	0	1	0	0	48	0	1
0	0	0	0	1	33	0	1
1	0	0	0	0	39	1	1
0	0	0	1	0	47	1	0

Pay Gap Decision Boundary

predict **1** when:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 \mathbf{>= 0}$$

predict **0** when:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 \mathbf{< 0}$$

Decision Boundary

$$\text{Pay Gap} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Decision Boundary

$$\text{Pay Gap} = -6 + 1x_1 + 1x_2$$

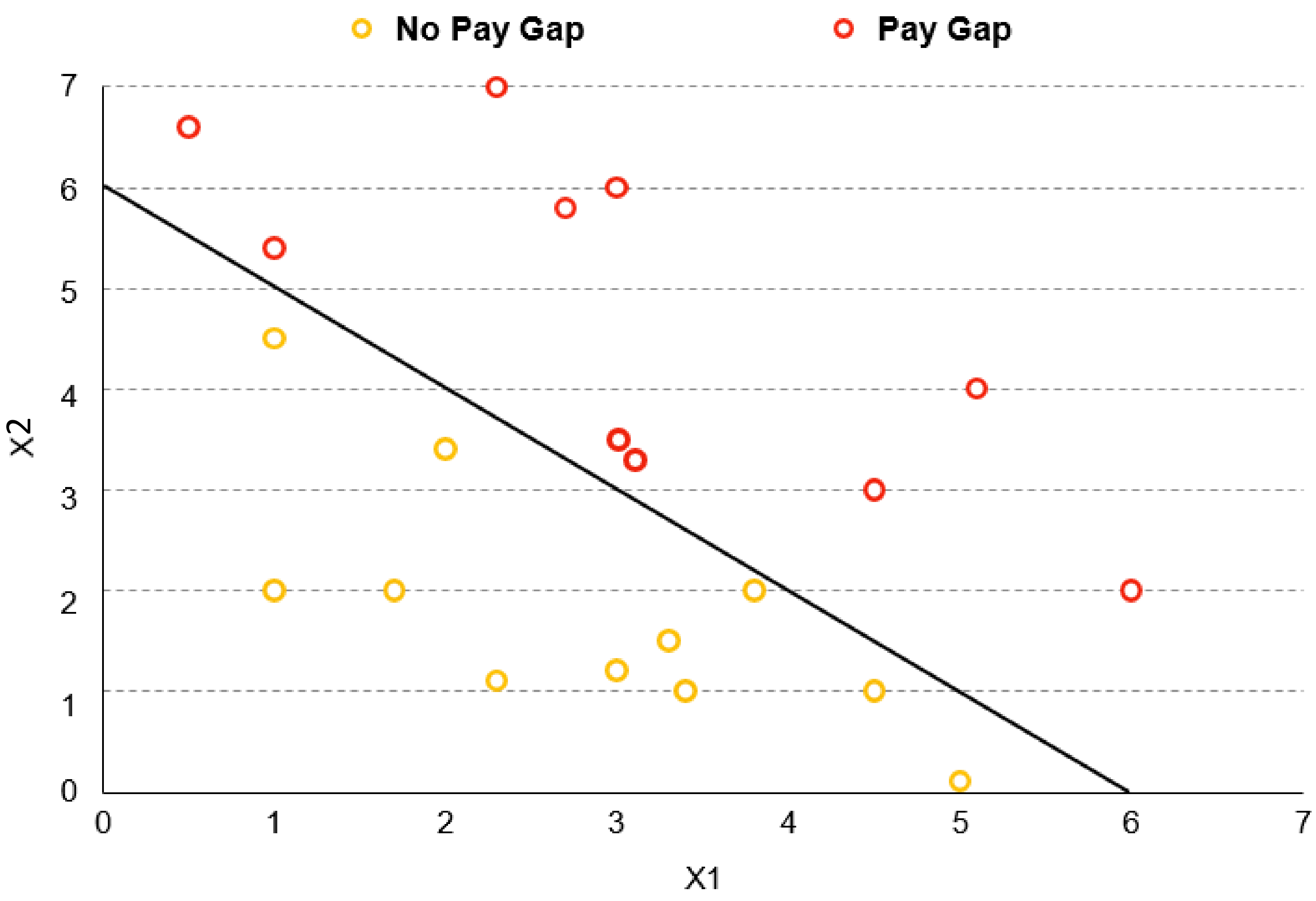
Decision Boundary

$$\text{Pay Gap} = -6 + 1x_1 + 1x_2$$

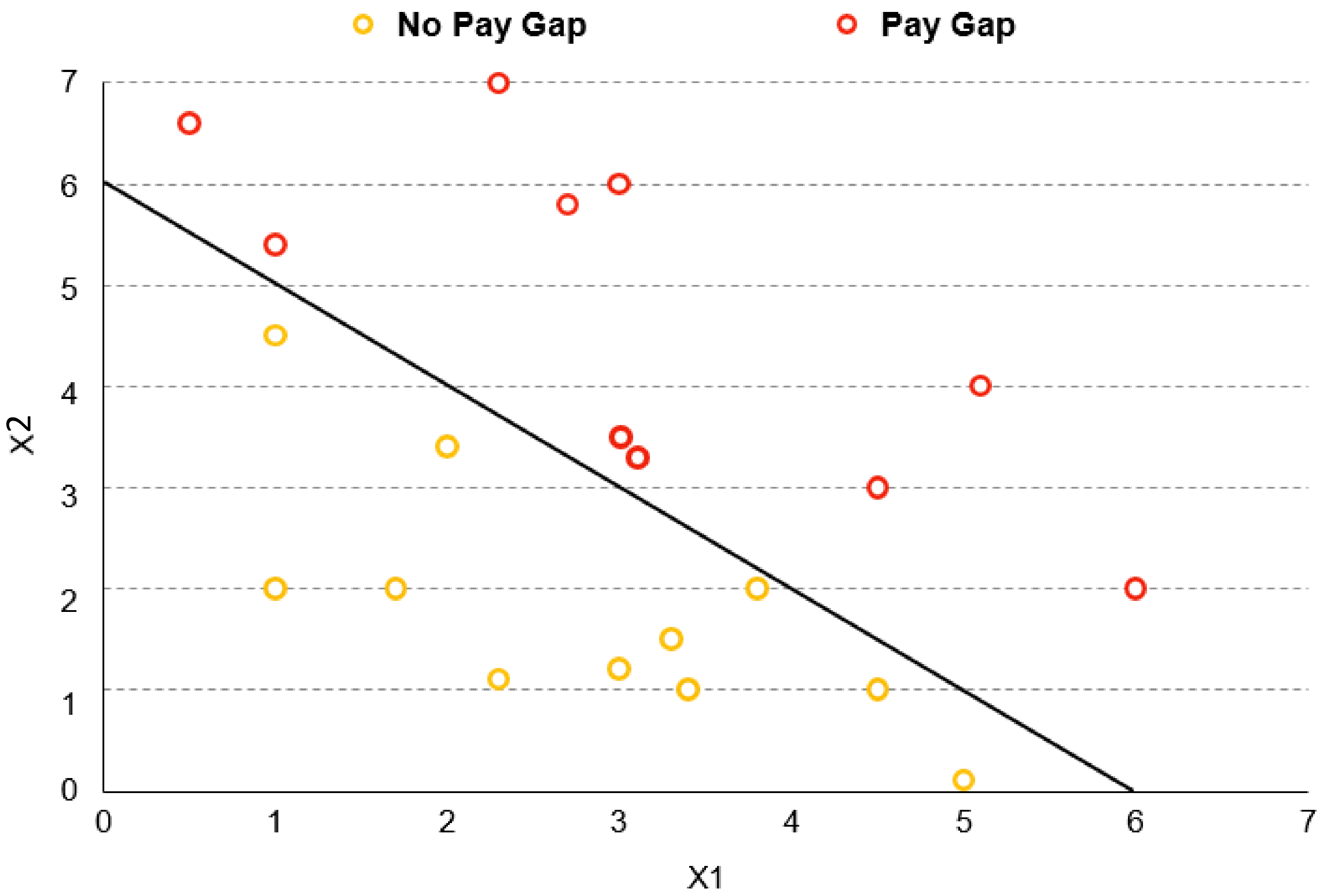
predict 1 when $-6 + 1x_1 + 1x_2 >= 0$

predict 0 when $-6 + 1x_1 + 1x_2 < 0$

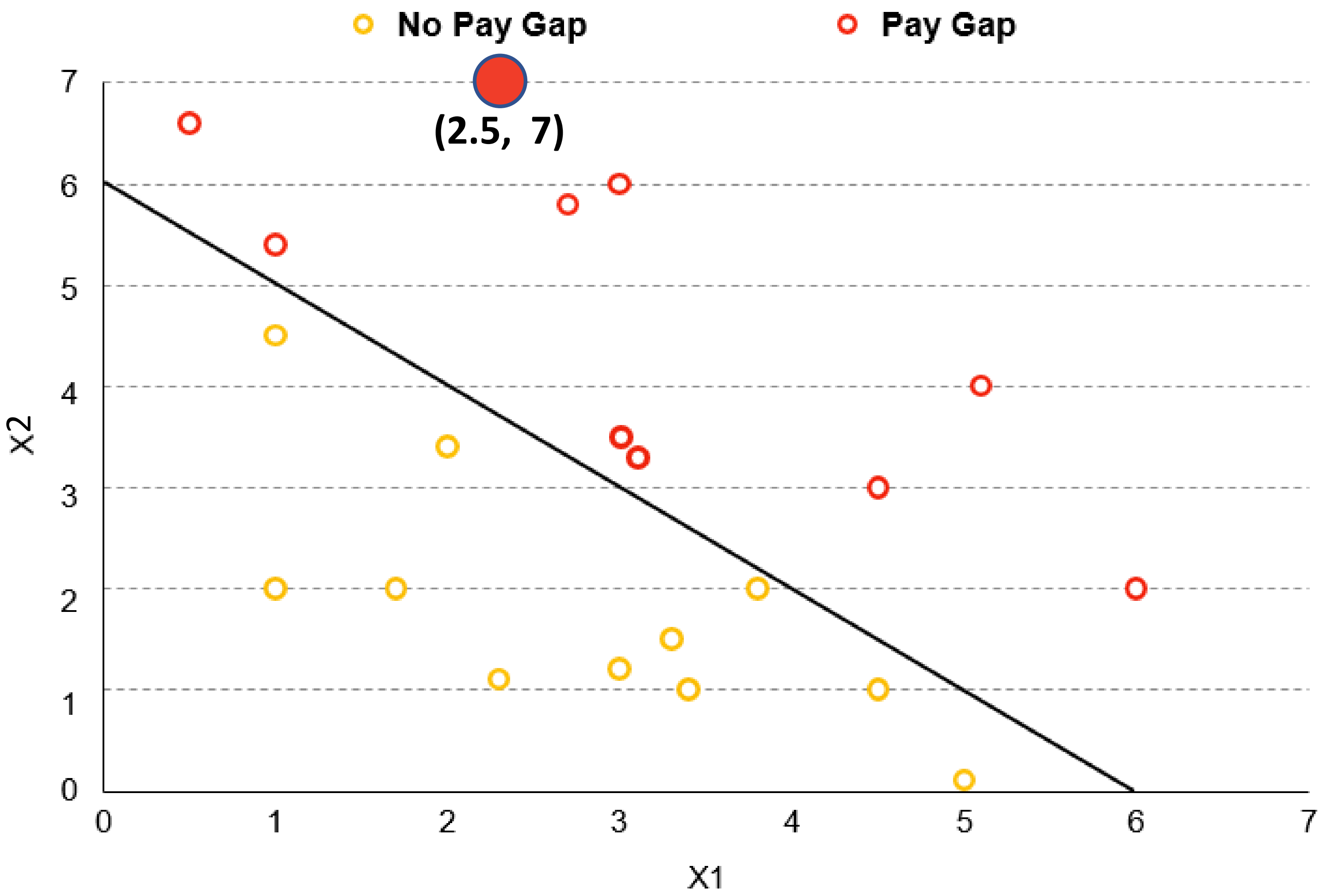
Pay Gap Decision Boundary



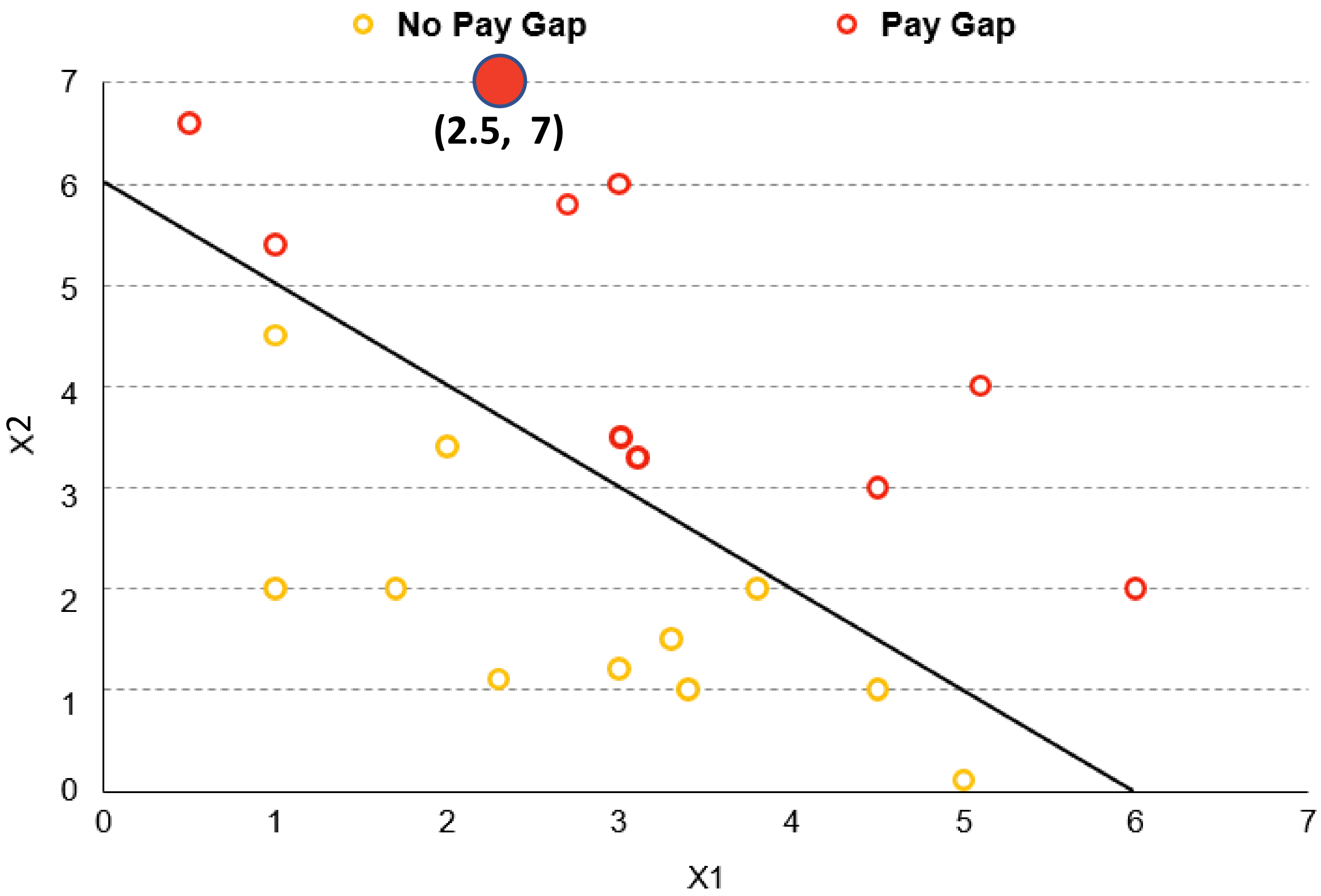
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 1x_1 + 1x_2)}}$$



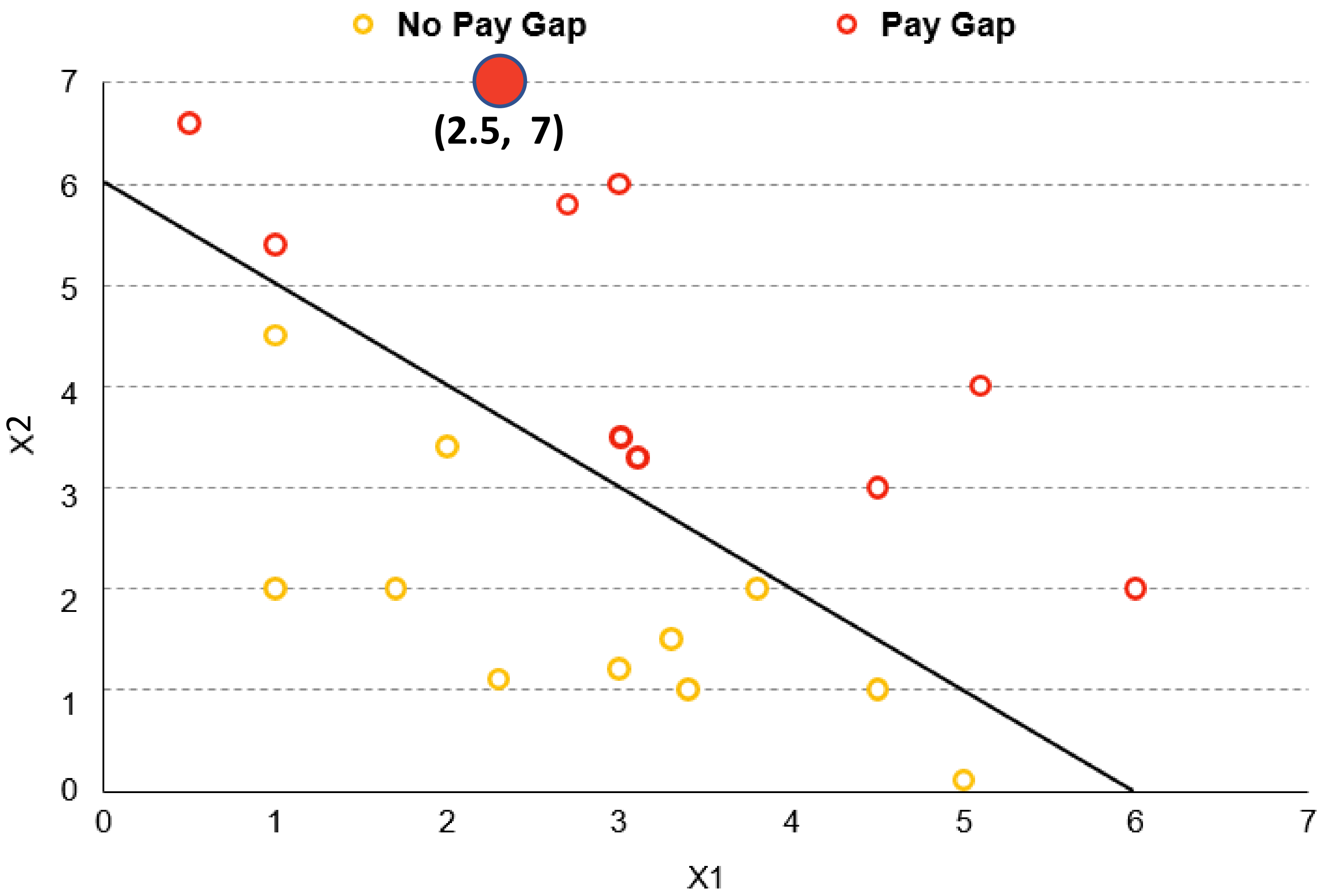
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 1x_1 + 1x_2)}}$$



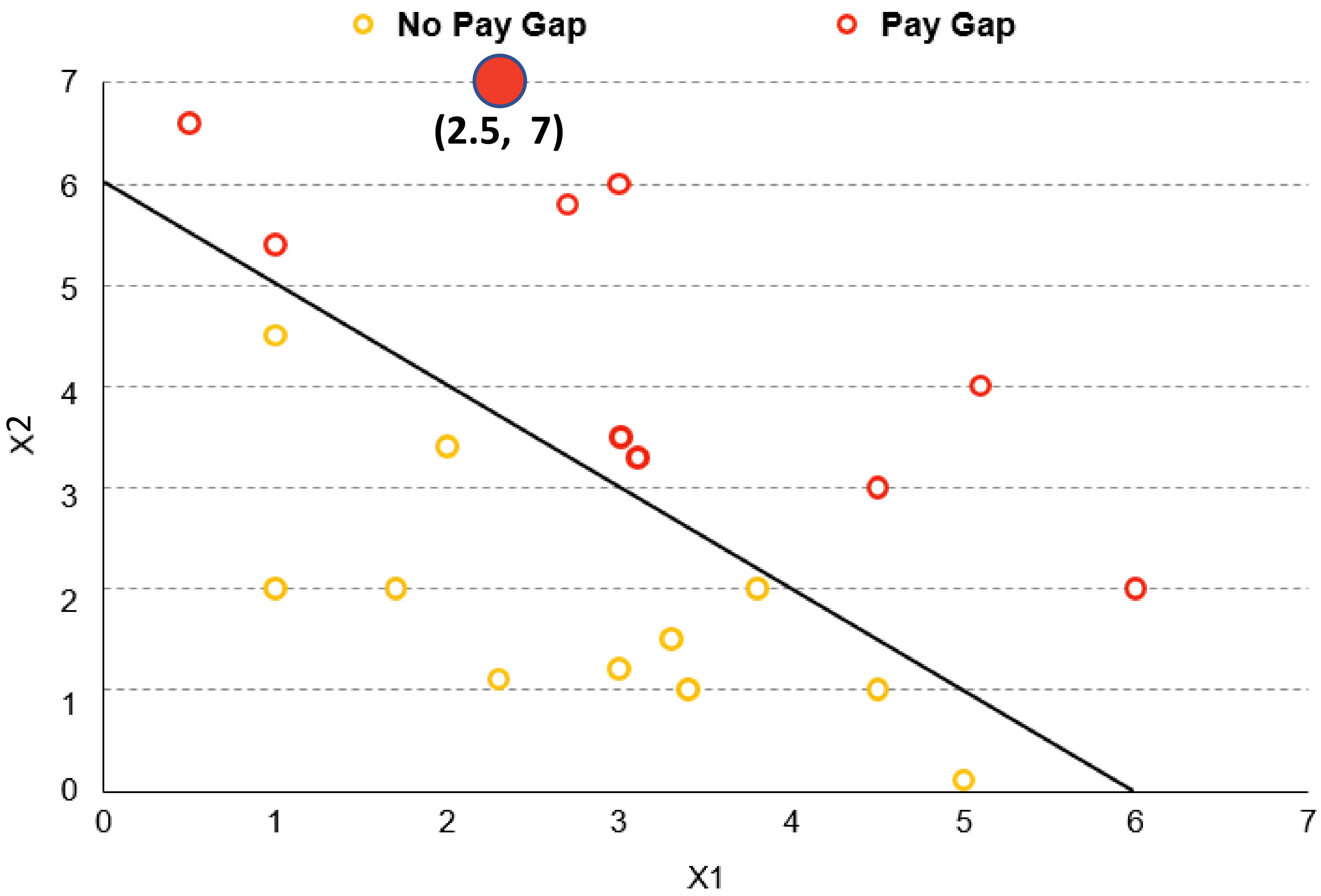
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 2.5 + 7)}}$$



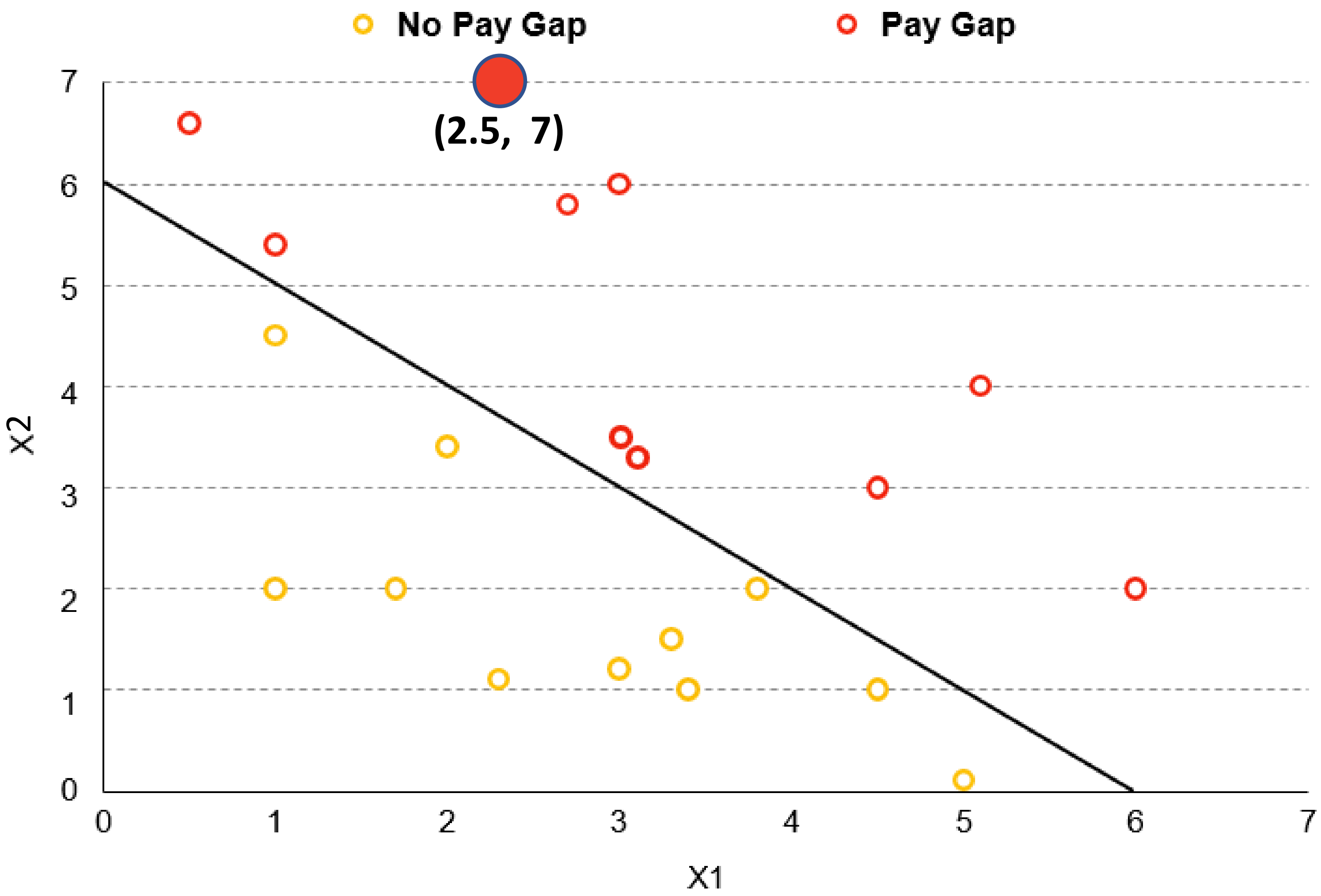
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 9.5)}}$$



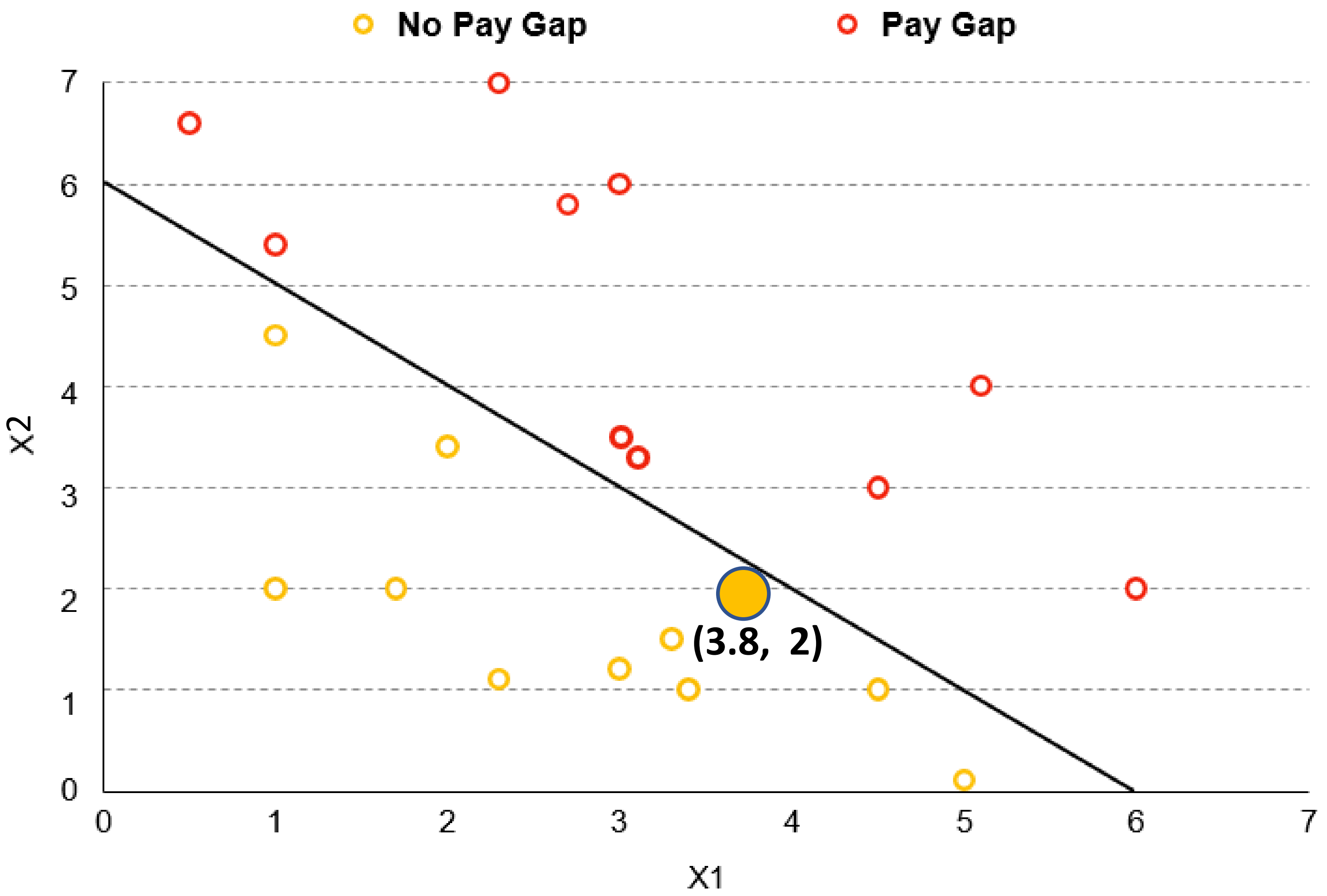
$$\hat{f}(X) = \frac{1}{1 + e^{-(3.5)}}$$



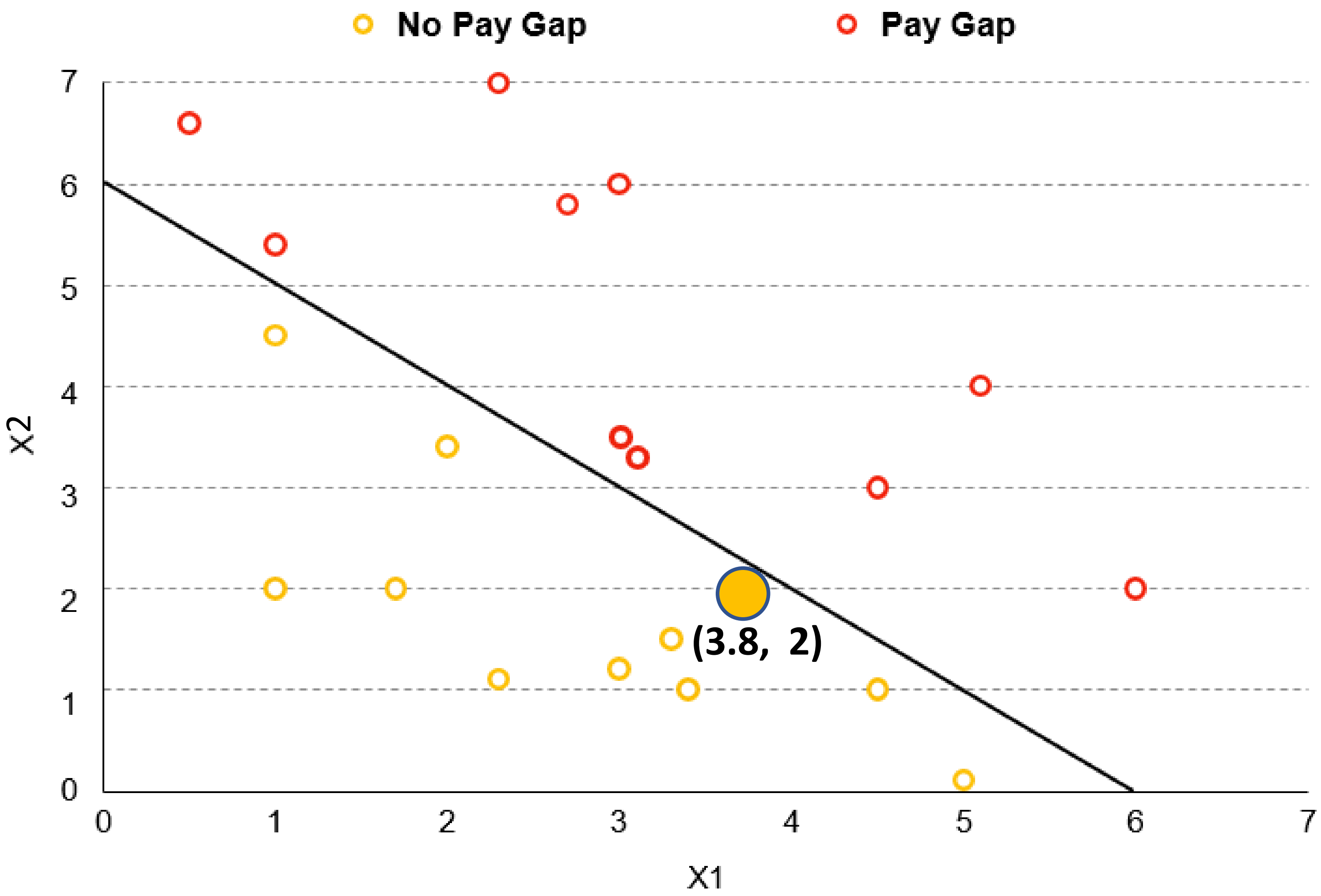
$$\hat{f}(X) = \frac{1}{1 + e^{-(3.5)}} = .97 \text{ (probability)}$$



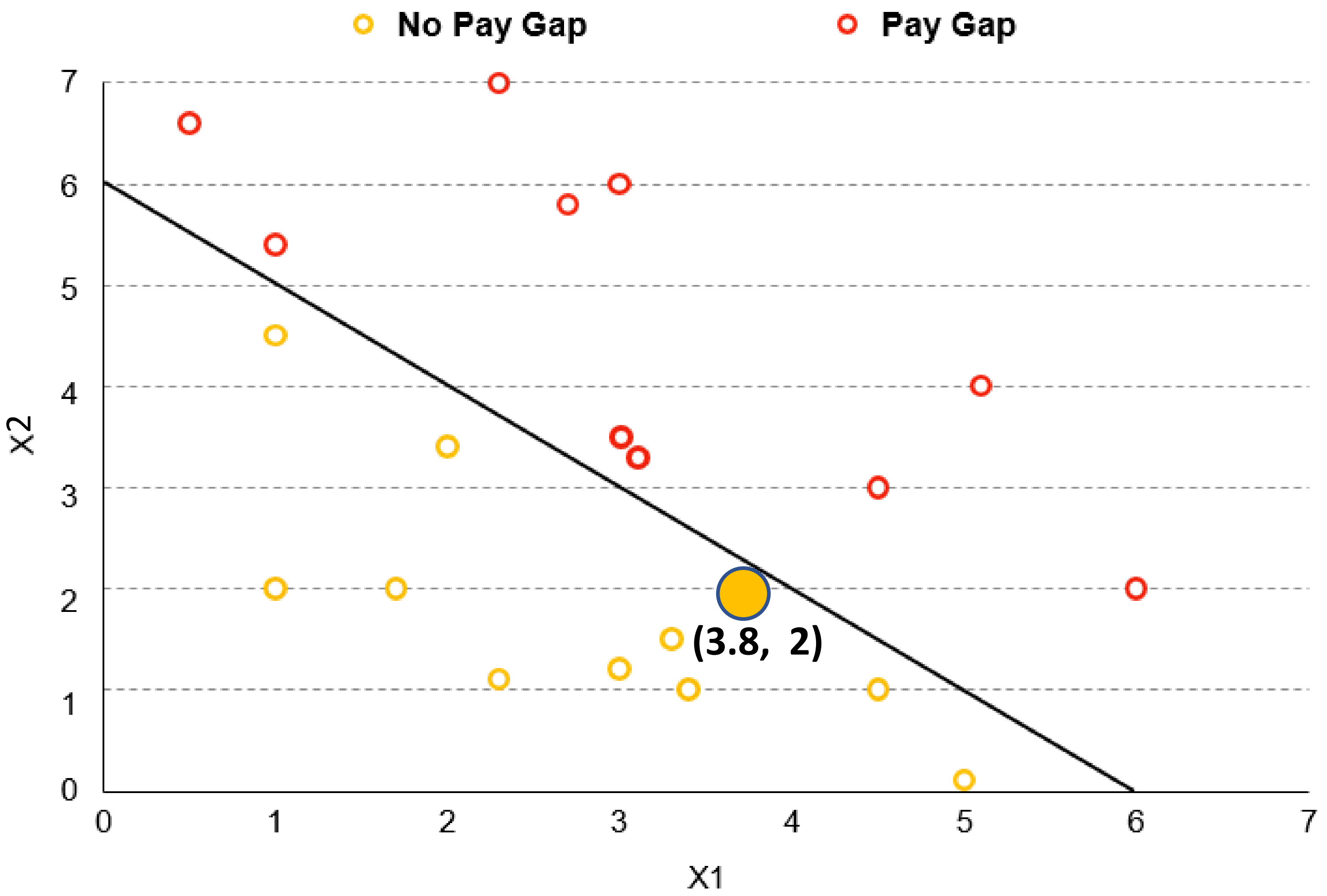
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 1x_1 + 1x_2)}}$$



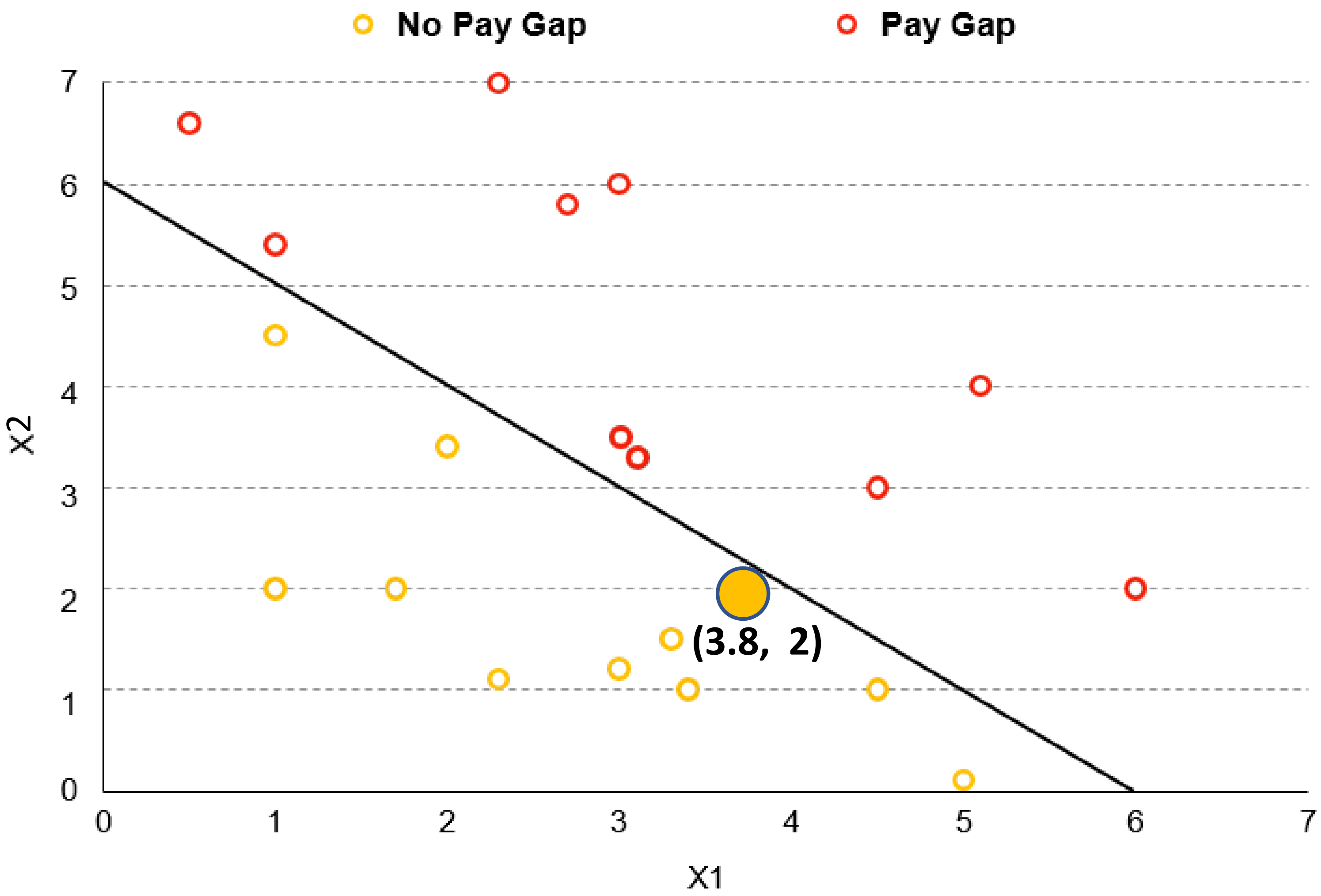
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 3.8 + 2)}}$$



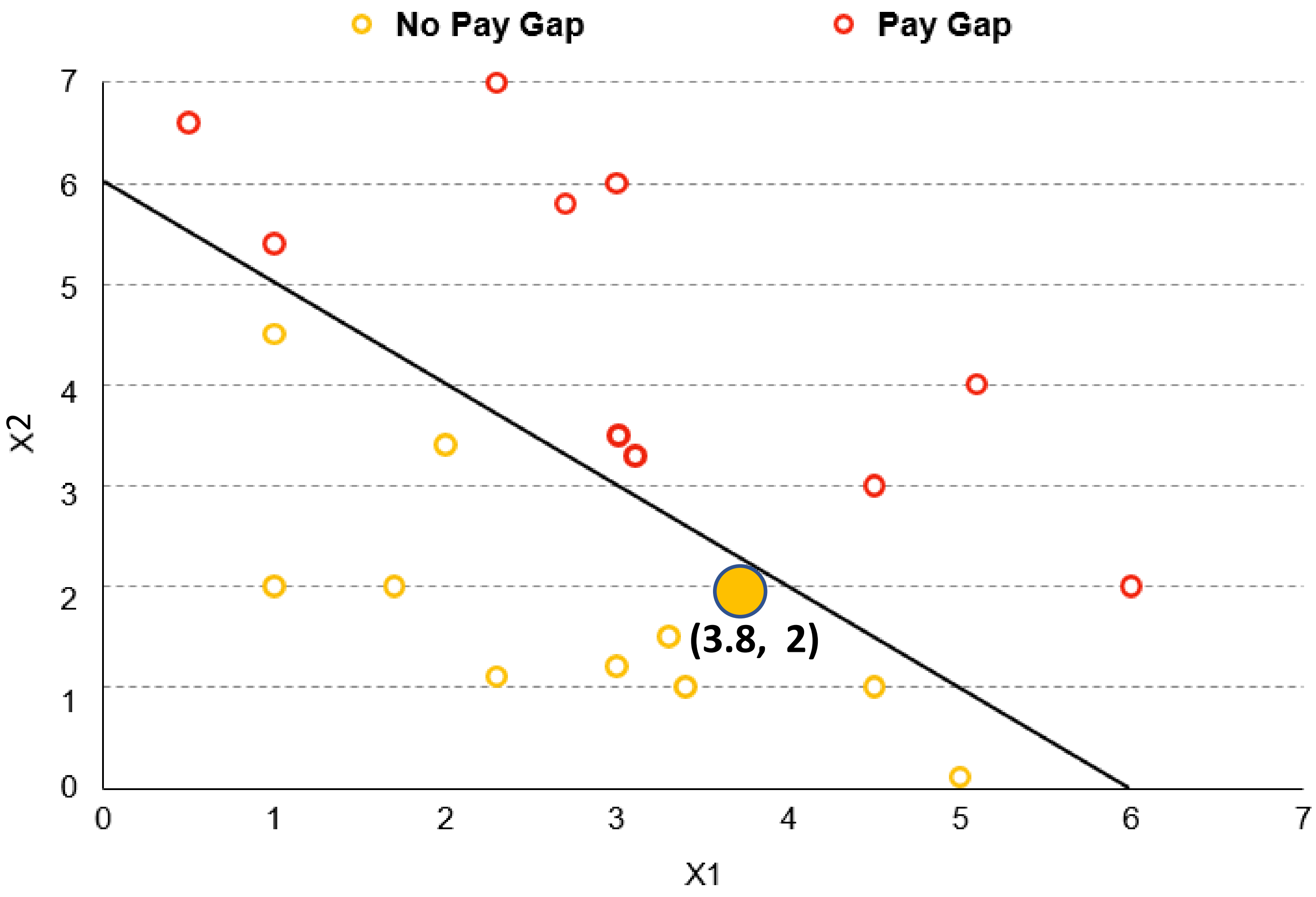
$$\hat{f}(X) = \frac{1}{1 + e^{-(-6 + 5.8)}}$$



$$\hat{f}(X) = \frac{1}{1 + e^{-(-0.2)}}$$



$$\hat{f}(X) = \frac{1}{1 + e^{-(-0.2)}} = .45 \text{ (probability)}$$



Support Vector Machine

Support Vector Machine

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

predict 1 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 >= 0$

predict 0 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 < 0$

Support Vector Machine

predict 1 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 >= 1$

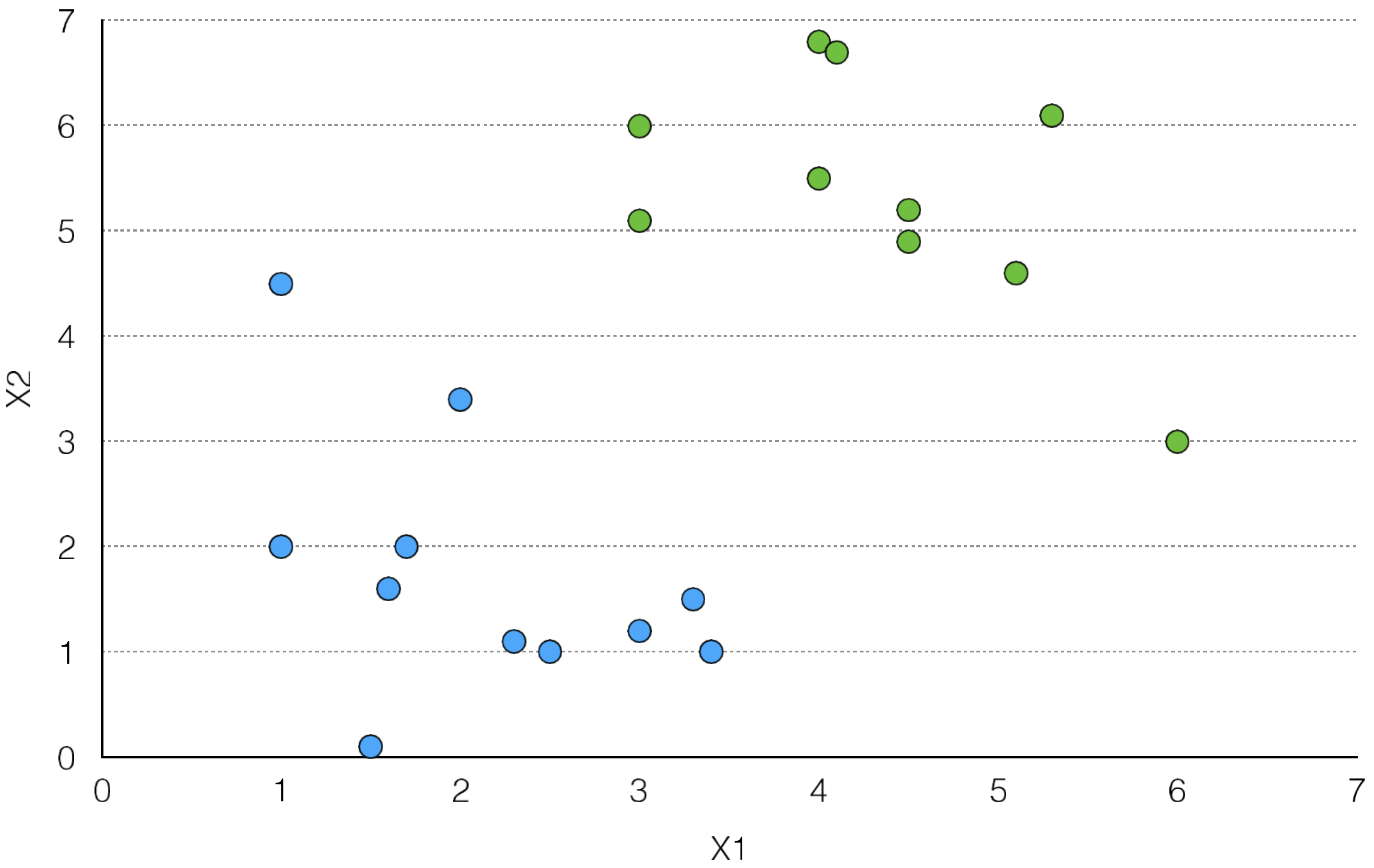
predict 0 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 < -1$

Support Vector Machine

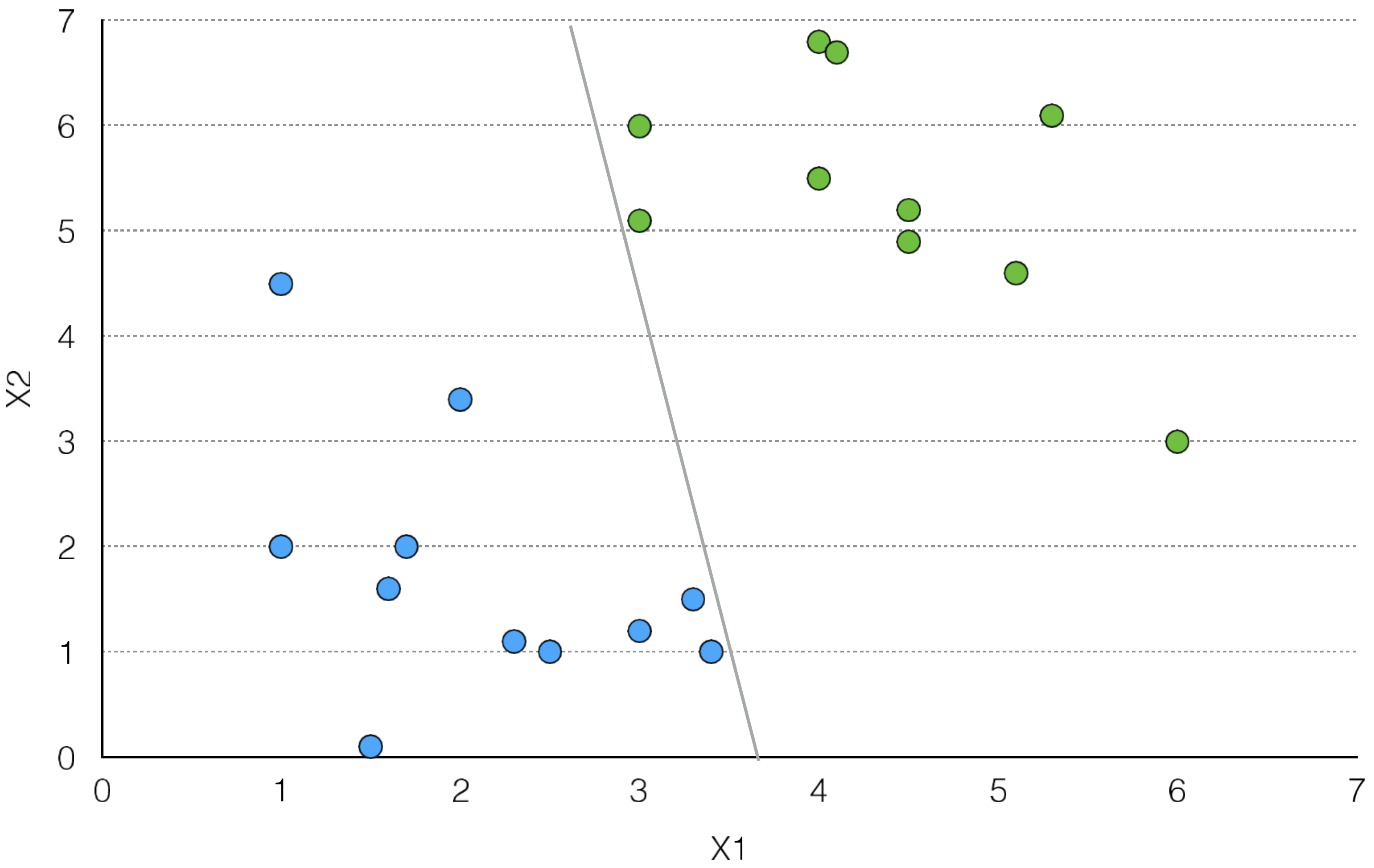
Large Margin Classifier

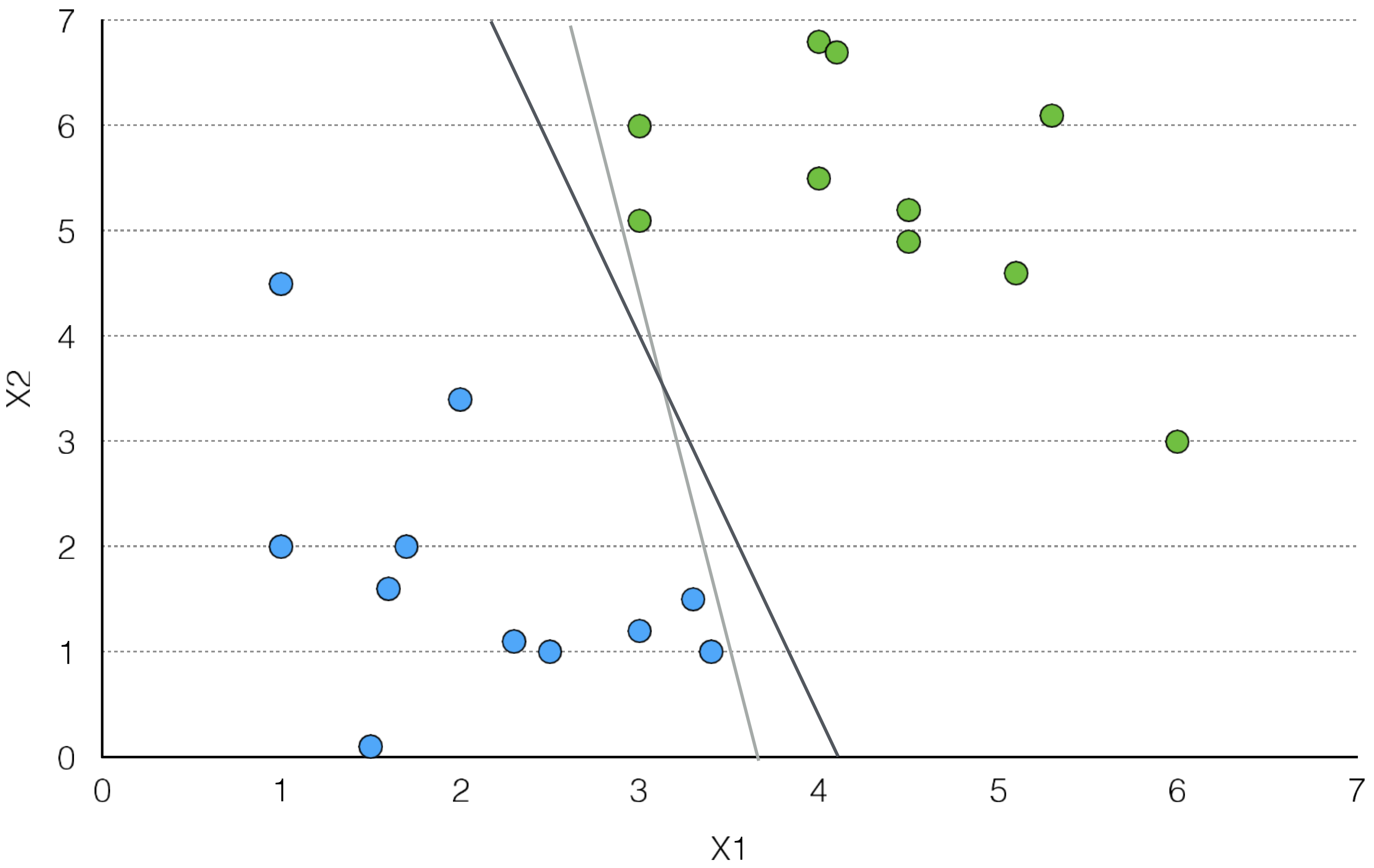
predict 1 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 >= 1$

predict 0 when $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 < -1$

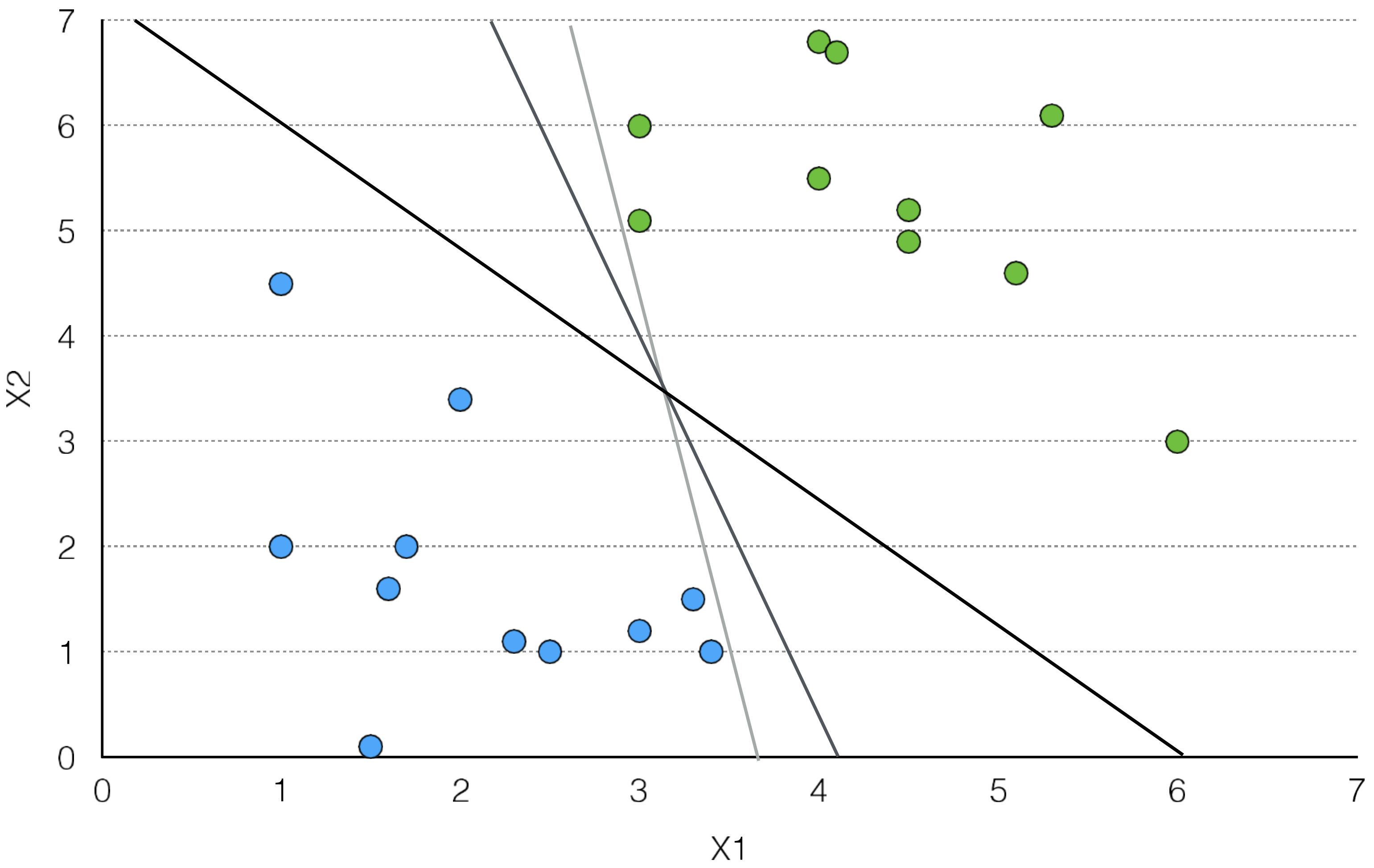


X2
X1
0
1
2
3
4
5
6
7

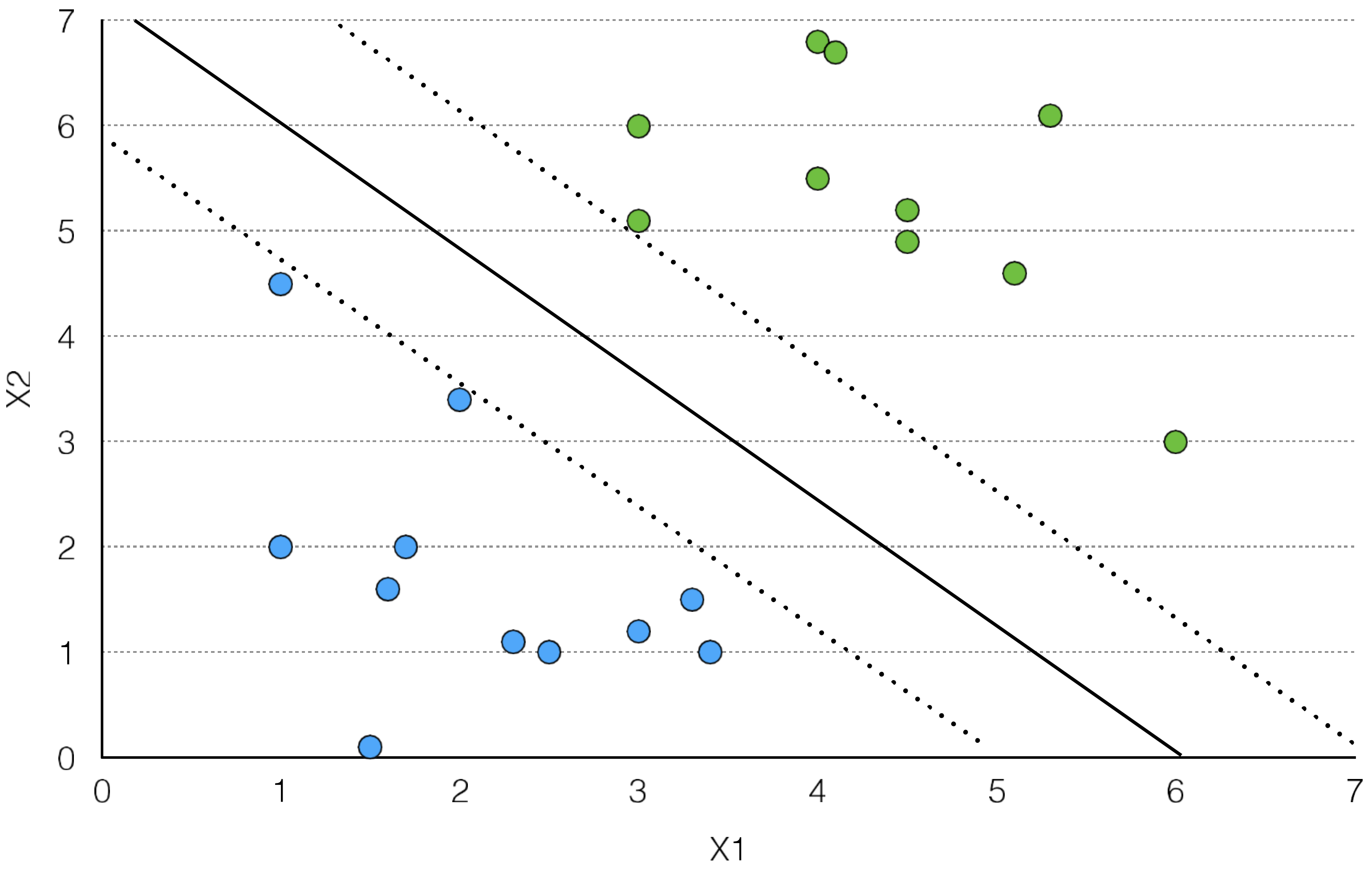

7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1



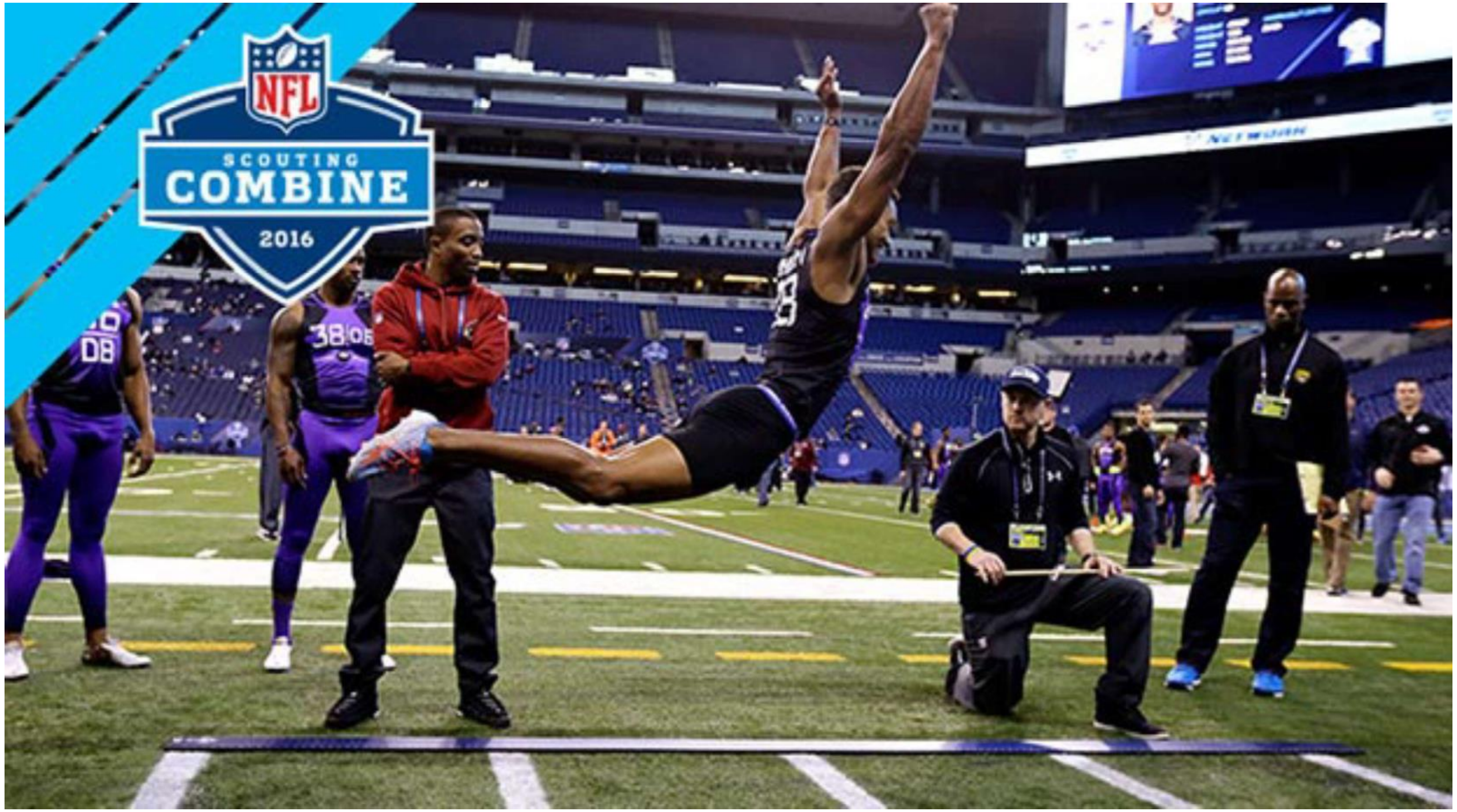
7
6
5
4
3
2
1
0
X2
0 1 2 3 4 5 6 7
X1



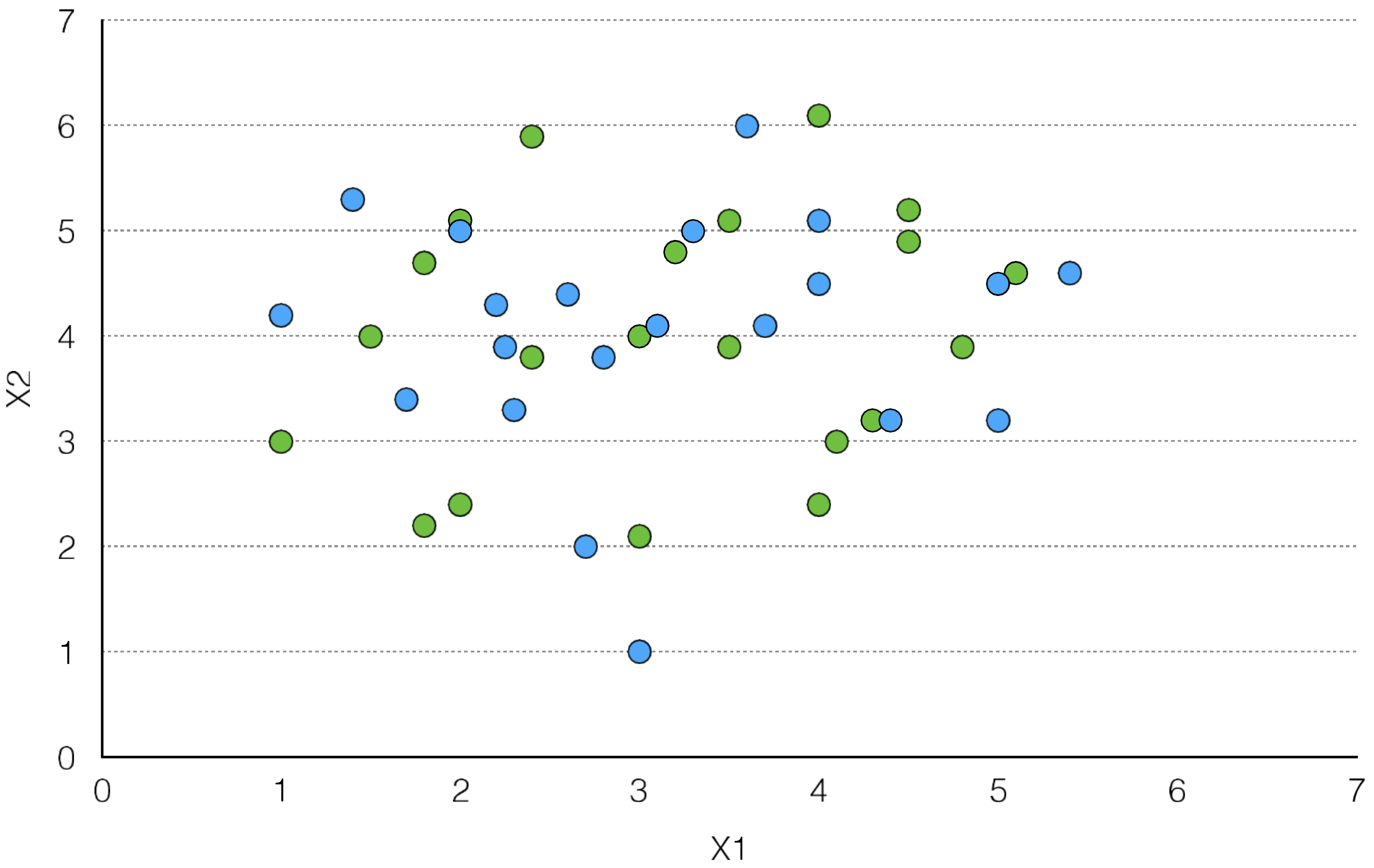
X2
X1
0
1
2
3
4
5
6
7



7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1


NFL
SCOUTING
COMBINE
2016

40-yard dash	Weight	Height	Drafted
5.10	290	74	1
4.92	275	75.5	1
4.43	178	69	0
4.62	221	74.5	1
4.91	248	75	0
5.53	303	77	0
4.47	189	71	1
4.56	205	71	1
4.75	267	73	0
4.84	261	74	1



7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1

40-yard dash	Weight	Height	Drafted
5.10	290	74	1
4.92	275	75.5	1
4.43	178	69	0
4.62	221	74.5	1
4.91	248	75	0
5.53	303	77	0
4.47	189	71	1
4.56	205	71	1
4.75	267	73	0
4.84	261	74	1

Feature Engineering

40-yard dash	BMI (wt/ht^2)	Drafted
5.10	37.2	1
4.92	33.9	1
4.43	26.3	0
4.62	28	1
4.91	31	0
5.53	35.9	0
4.47	26.4	1
4.56	28.6	1
4.75	35.2	0
4.84	33.5	1

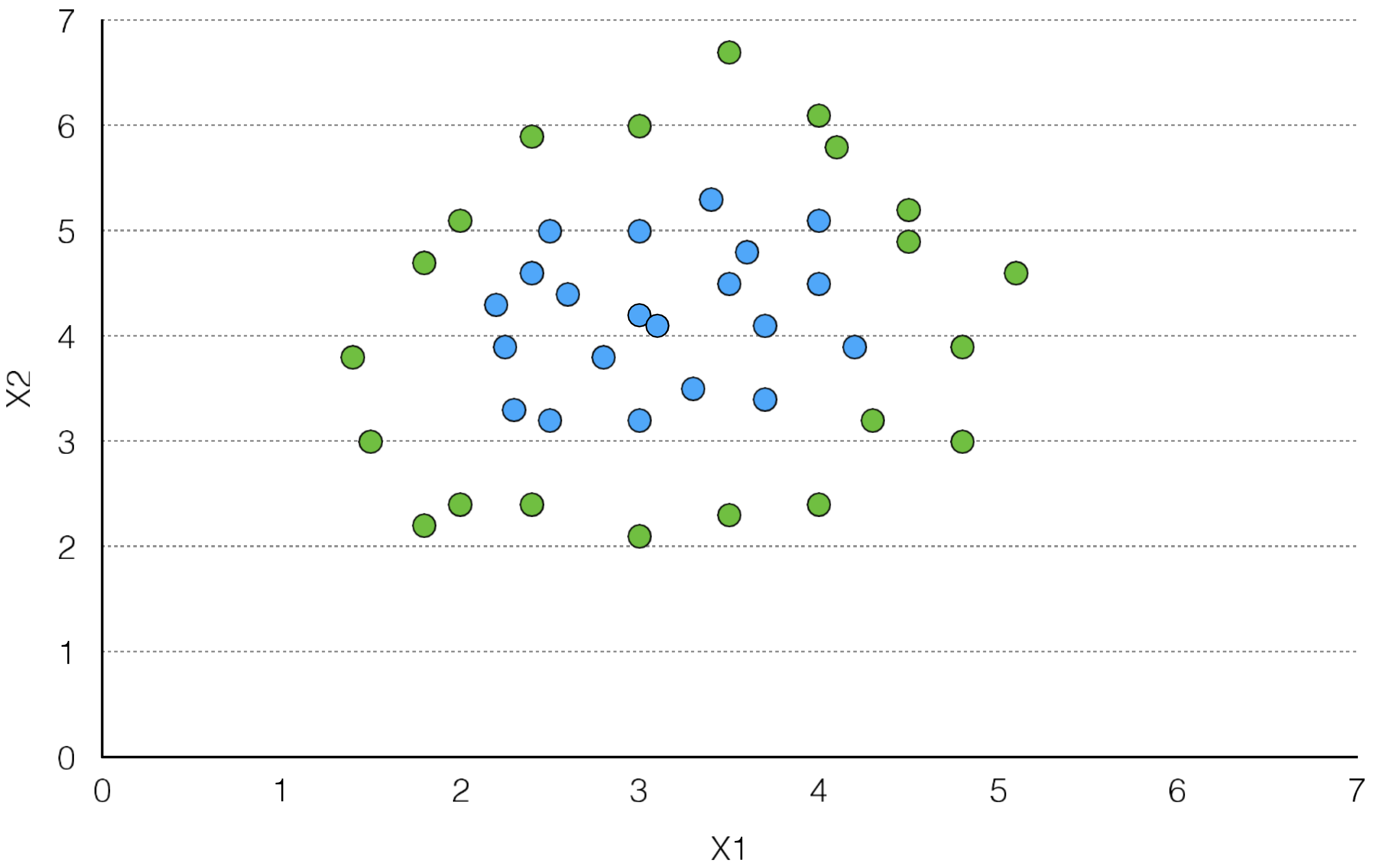

44
OL

Feature Engineering

40-yard dash	BMI (wt/ht^2)	Drafted
5.10	37.2	1
4.92	33.9	1
4.43	26.3	0
4.62	28	1
4.91	31	0
5.53	35.9	0
4.47	26.4	1
4.56	28.6	1
4.75	35.2	0
4.84	33.5	1

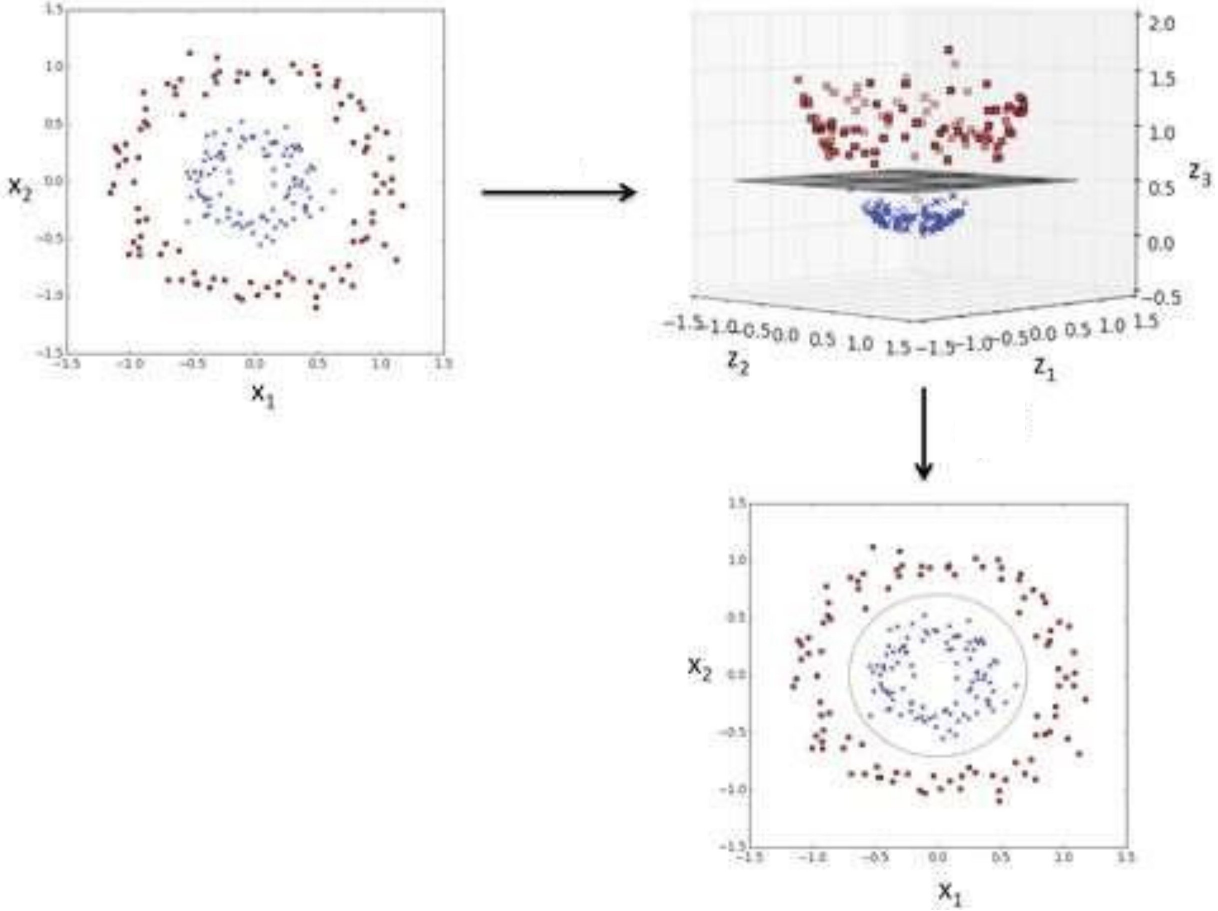
Feature Engineering

Speed-to-Size (40-yd/bsa)	BMI (wt/ht^2)	Drafted
2.16	37.2	1
2.06	33.9	1
2.02	26.3	0
1.97	28	1
2.23	31	0
2.00	35.9	0
2.03	26.4	1
1.99	28.6	1
1.85	35.2	0
2.03	33.5	1

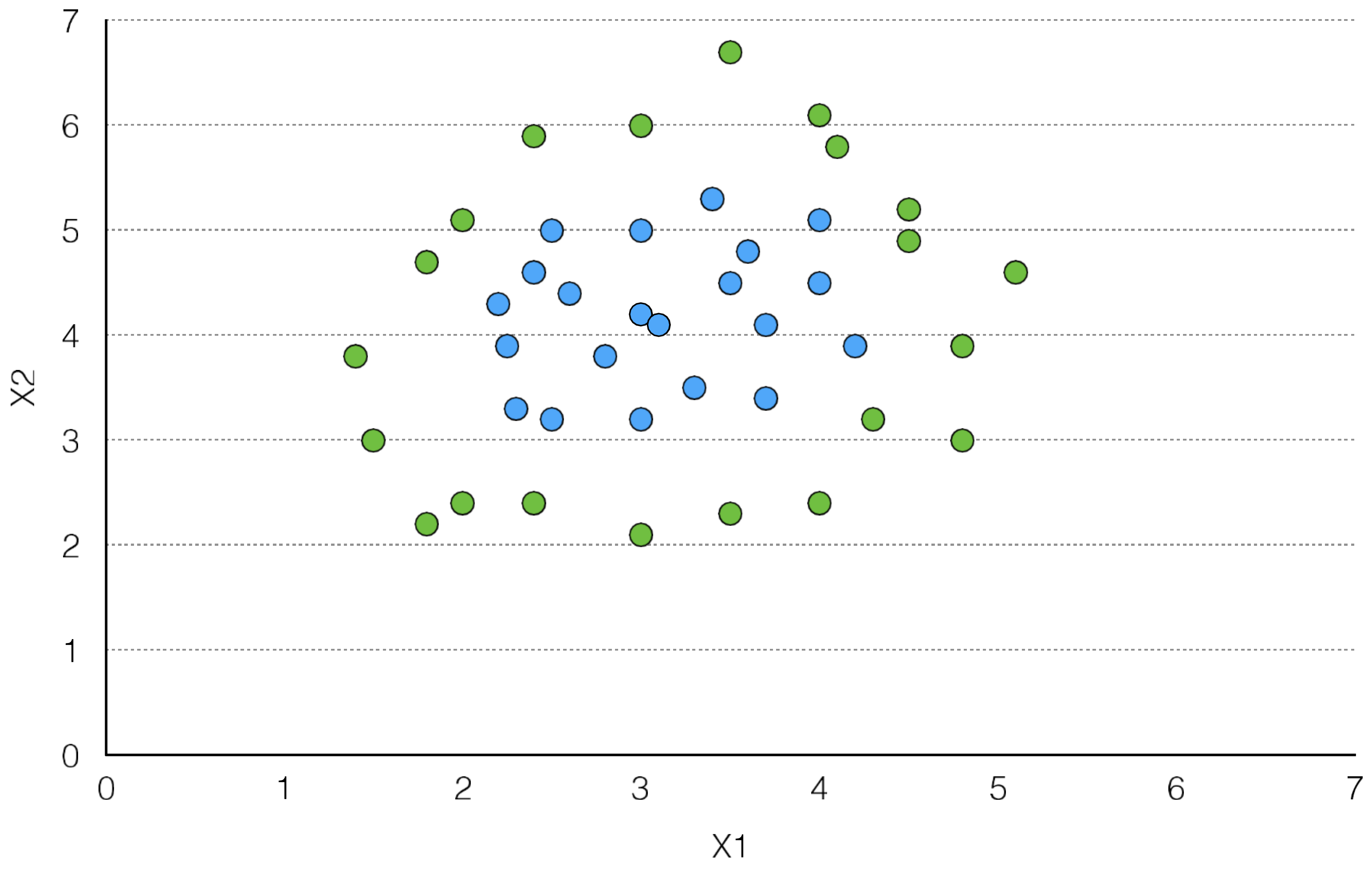

7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1

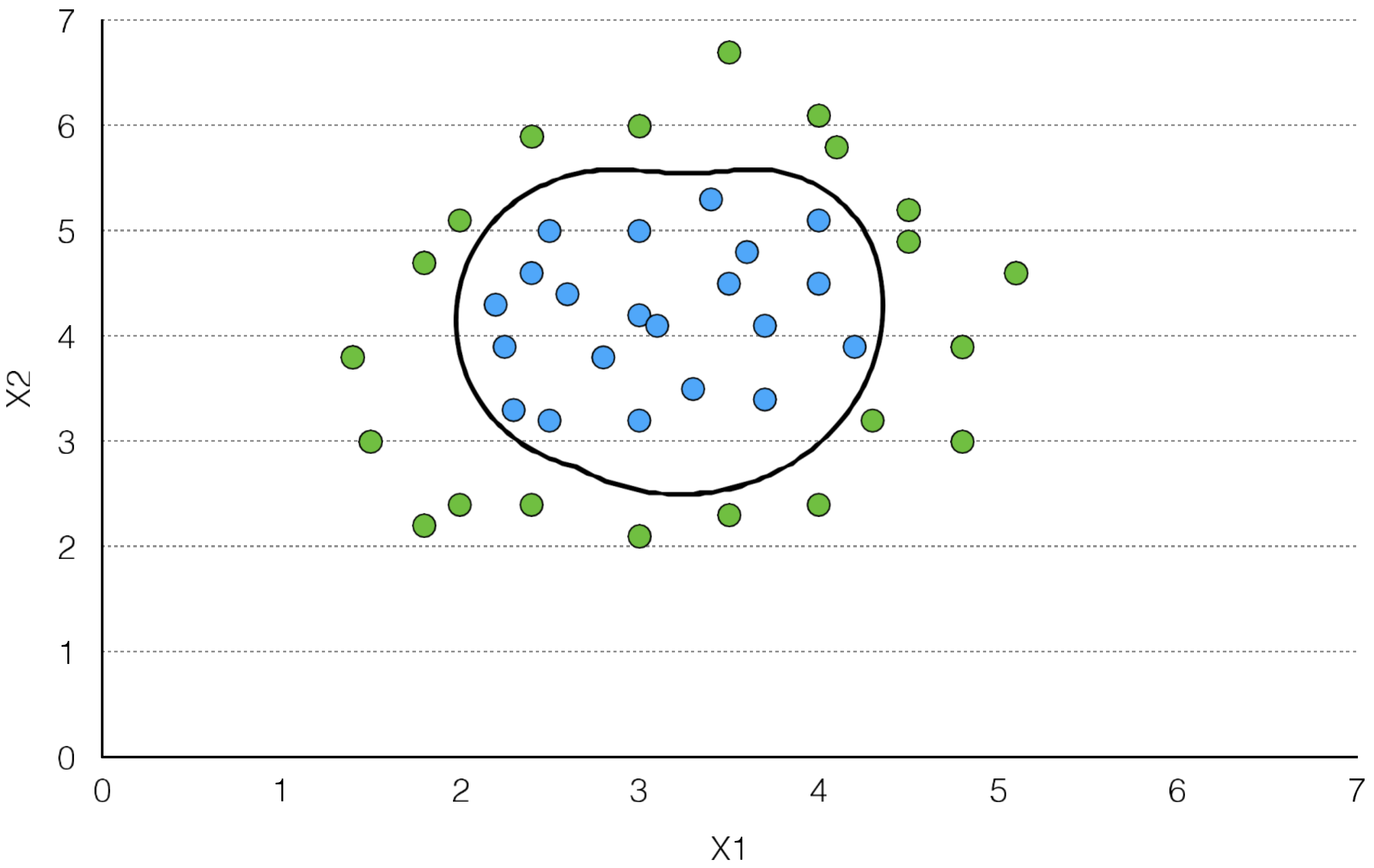
Kernel

Non-linear Classification



x_2
x_1
z_3
z_2
z_1
2.0
1.5
1.0
0.5
0.0
−0.5
x_2
x_1

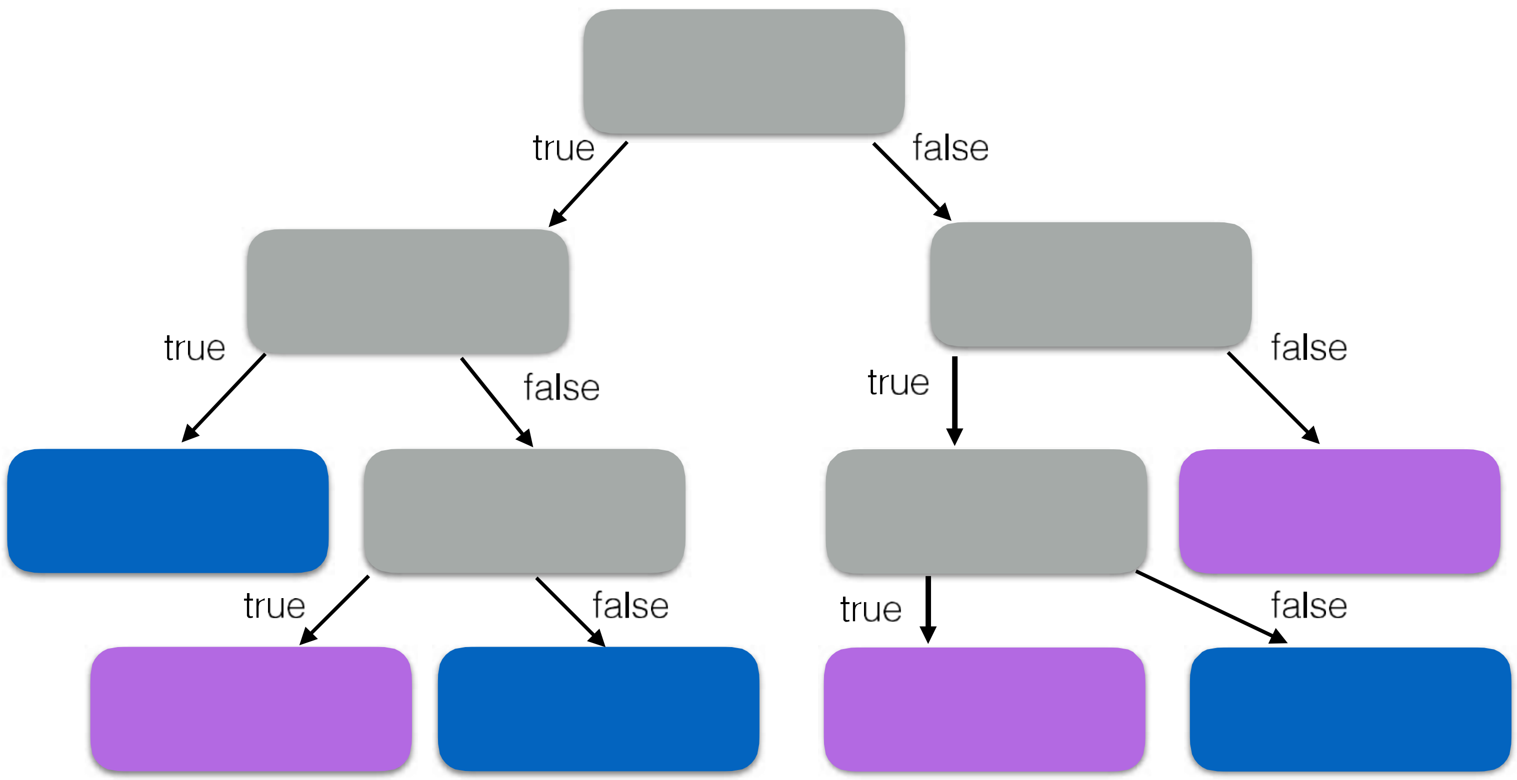

7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1



7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1

Decision Tree

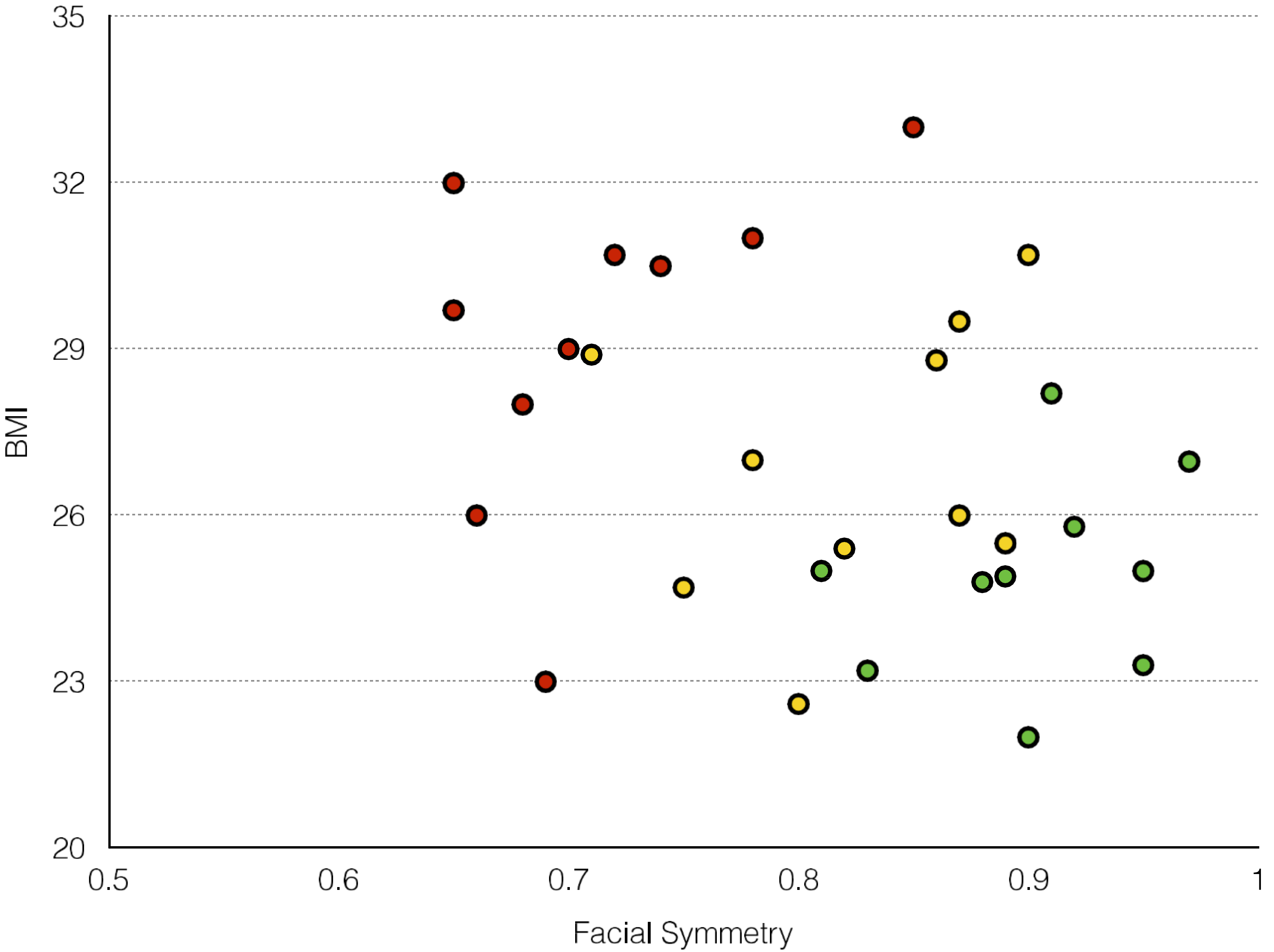
Decision Tree

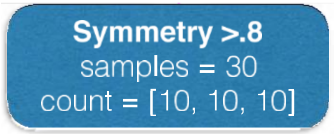


Short-term Attractiveness

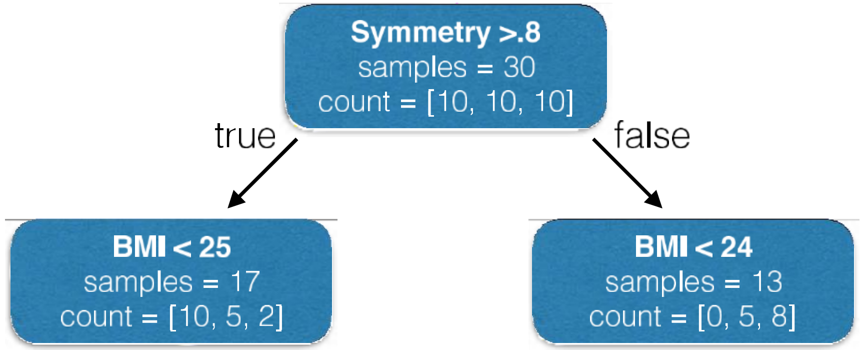
Short-term Attractiveness

Facial Symmetry	BMI	Waist-to-Hip	Well-Groomed
0.9	23.4	0.93	1
0.85	27.9	0.87	0
0.65	27.1	0.79	1
0.85	22.6	0.91	1
0.9	30.3	0.82	0
0.75	29.0	0.82	0
0.85	22.3	0.89	1
0.7	37.6	0.73	0
0.85	24.2	0.85	0

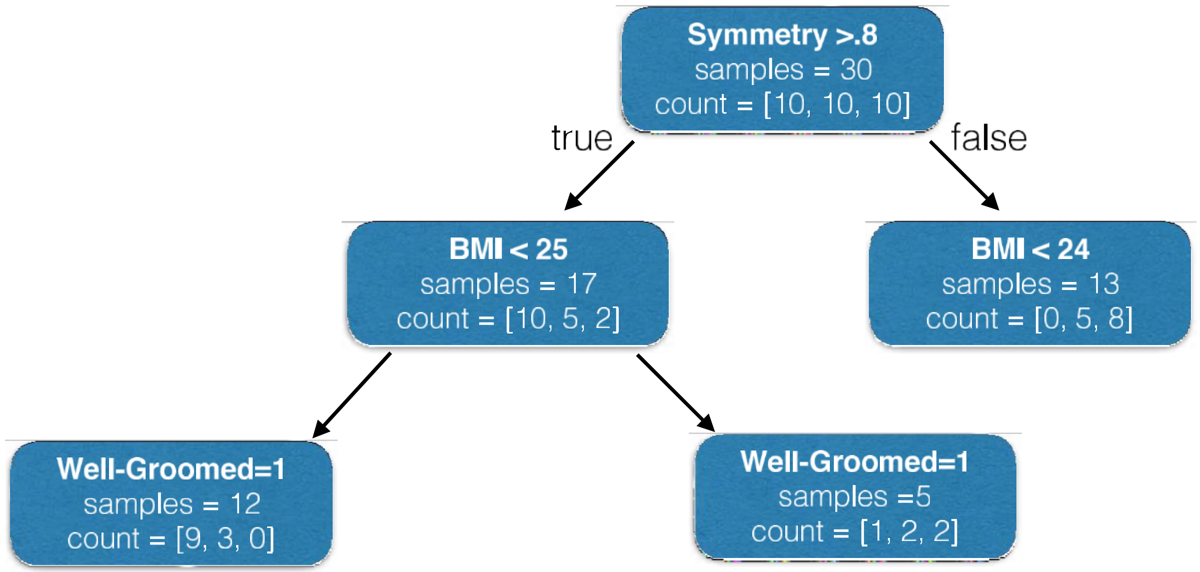

35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry



[att, ave, un]

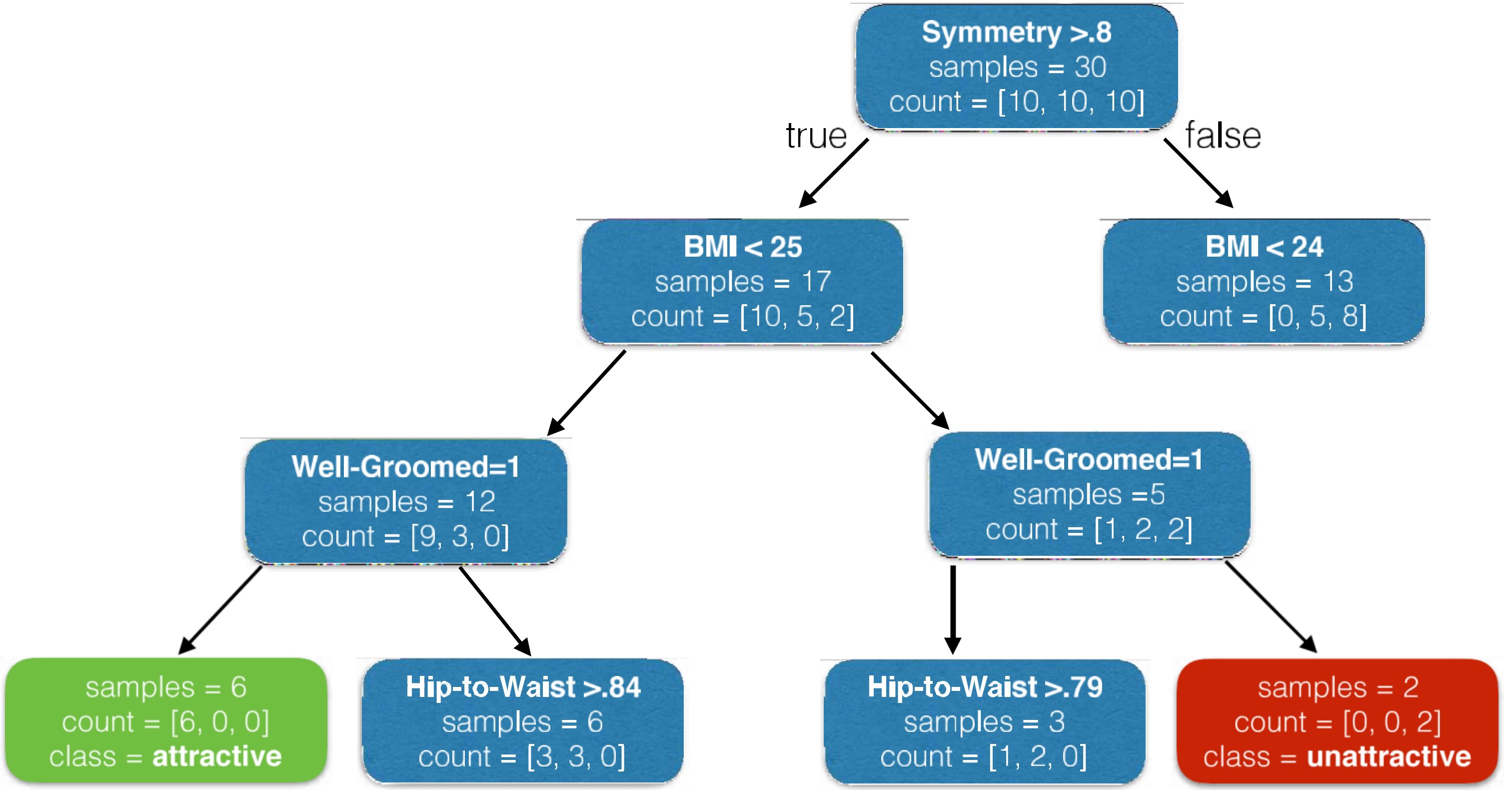

Symmetry >.8
samples = 30
count = [10, 10, 10]
true
false
BMI < 25
samples = 17
count = [10, 5, 2]
BMI < 24
samples = 13
count = [0, 5, 8]

[att, ave, un]

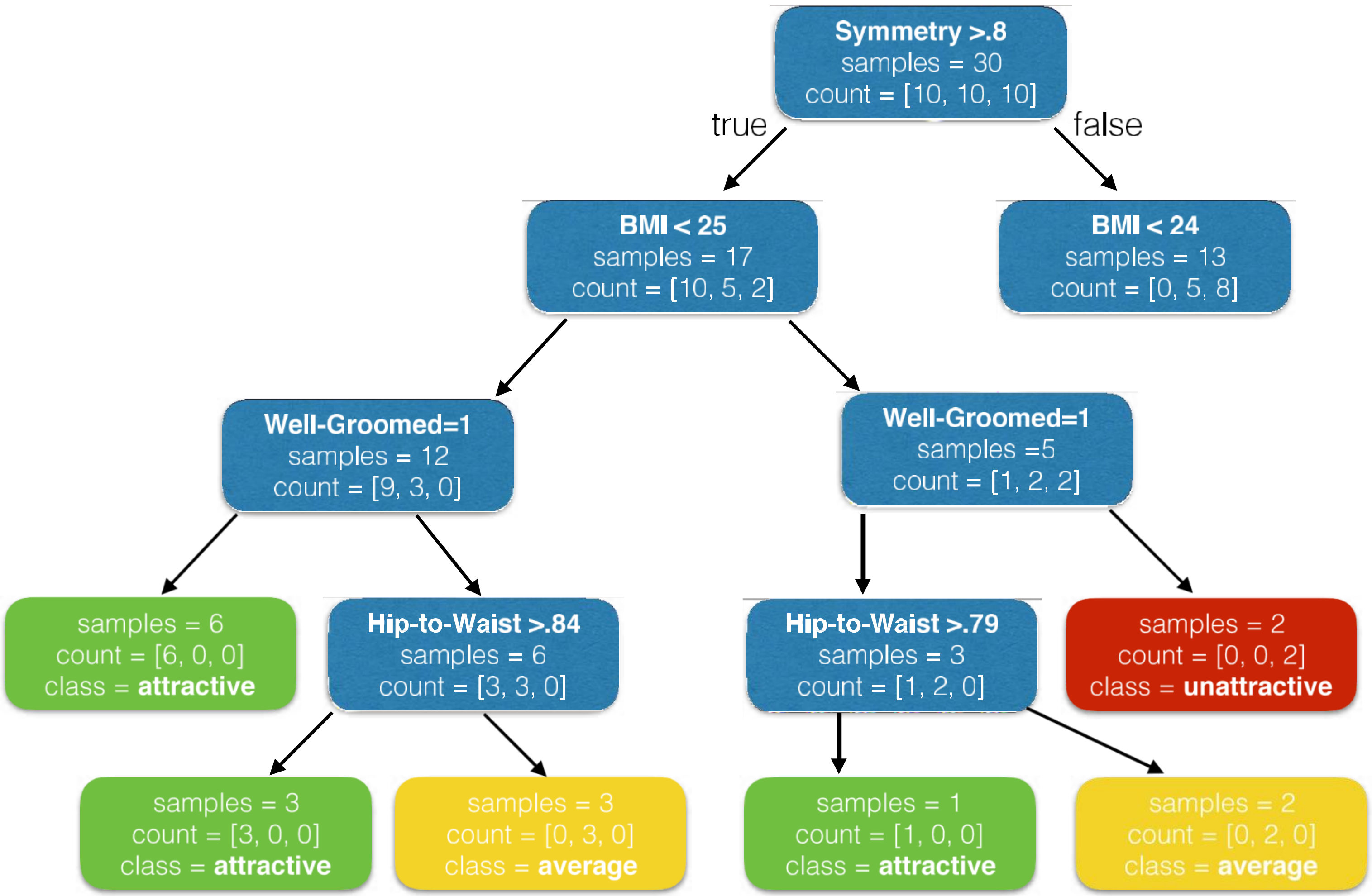


Symmetry >.8
samples = 30
count = [10, 10, 10]
true
false
BMI < 25
samples = 17
count = [10, 5, 2]
BMI < 24
samples = 13
count = [0, 5, 8]
Well-Groomed=1
samples = 12
count = [9, 3, 0]
Well-Groomed=1
samples =5
count = [1, 2, 2]

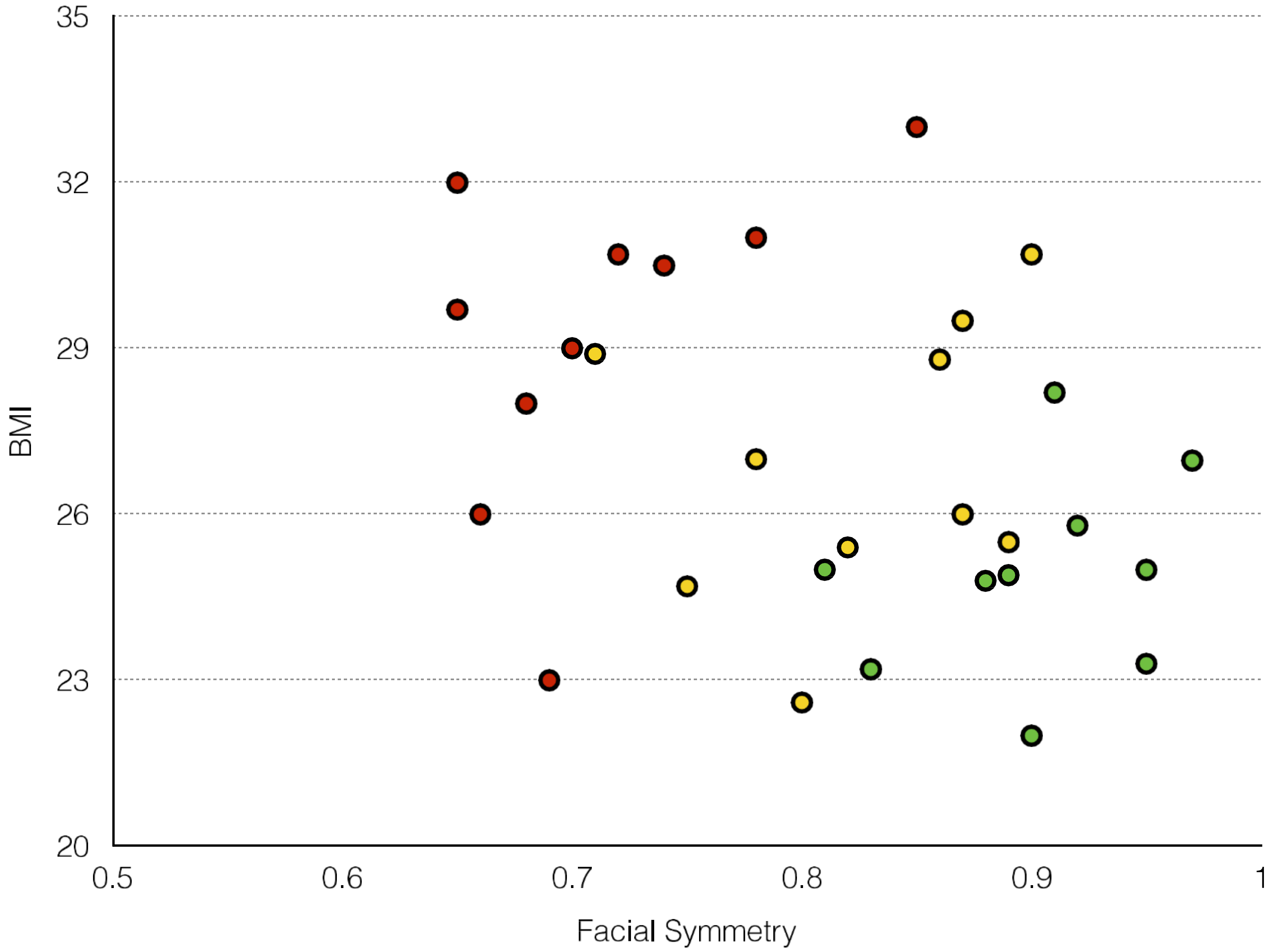
[att, ave, un]

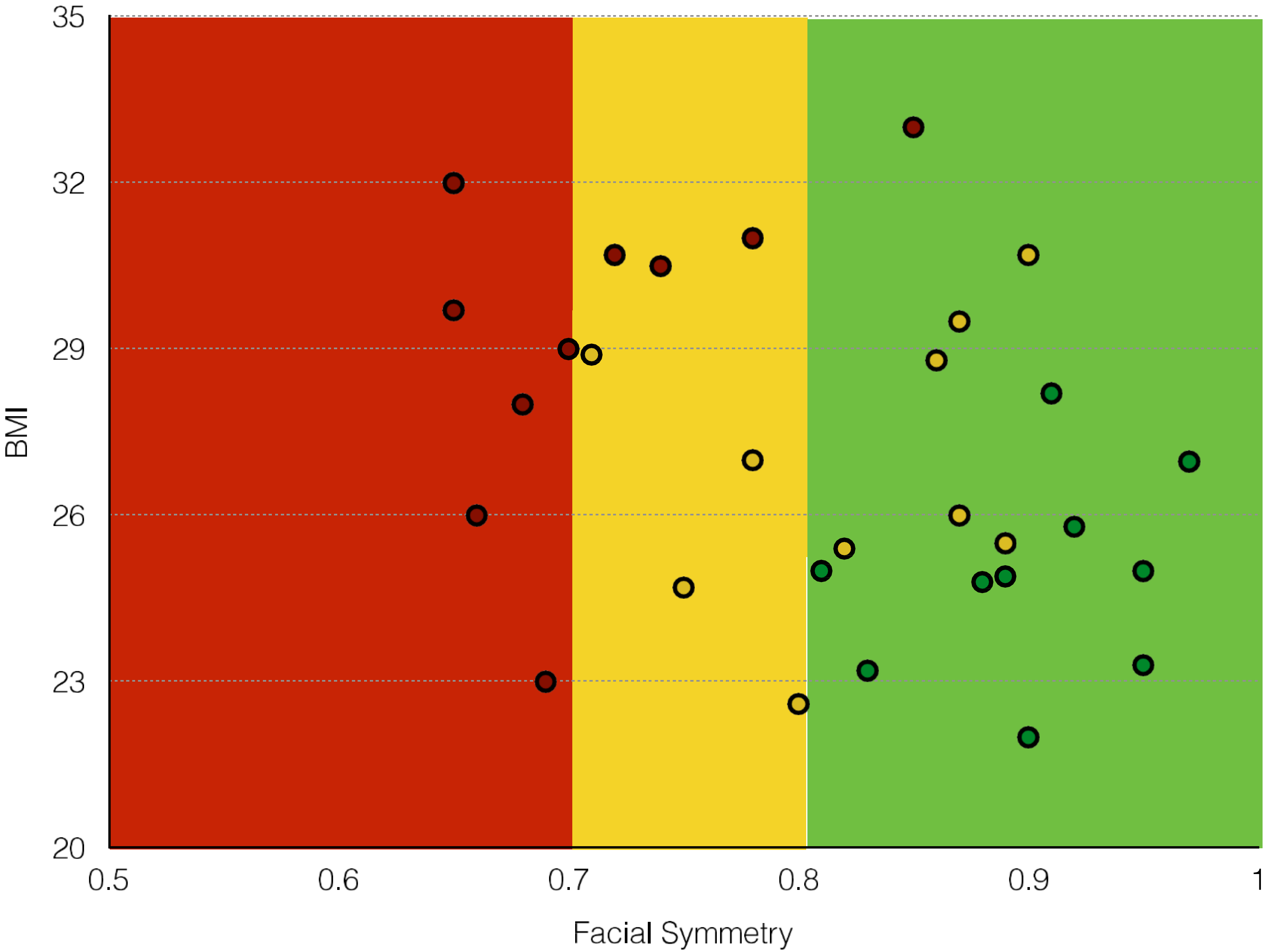

Symmetry >.8
samples = 30
count = [10, 10, 10]
true
false
BMI < 25
samples = 17
count = [10, 5, 2]
BMI < 24
samples = 13
count = [0, 5, 8]
Well-Groomed=1
samples = 12
count = [9, 3, 0]
Well-Groomed=1
samples =5
count = [1, 2, 2]
samples = 6
count = [6, 0, 0]
class = attractive
Hip-to-Waist >.84
samples = 6
count = [3, 3, 0]
Hip-to-Waist >.79
samples = 3
count = [1, 2, 0]
samples = 2
count = [0, 0, 2]
class = unattractive

[att, ave, un]

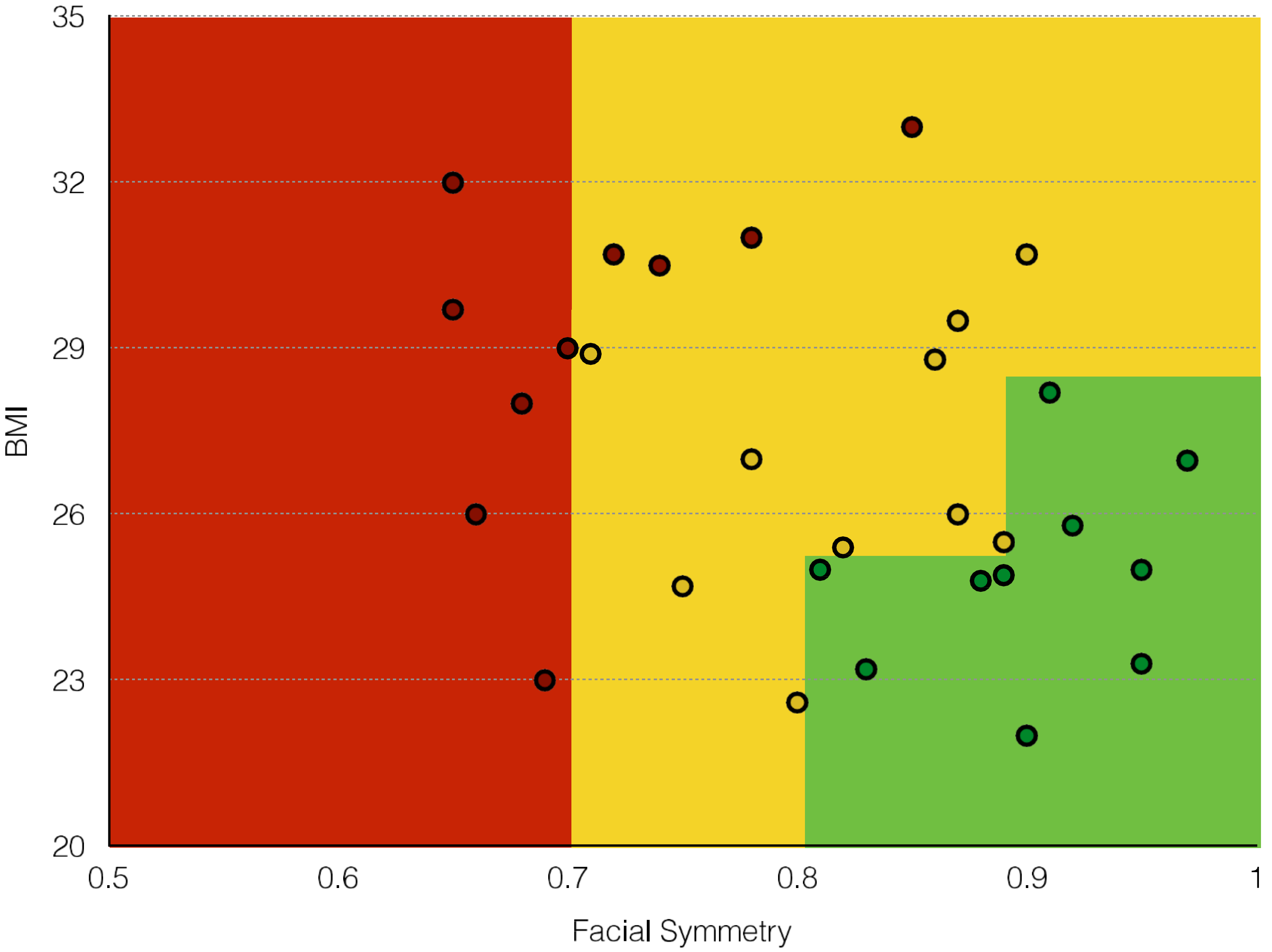

Symmetry >.8
samples = 30
count = [10, 10, 10]
true
false
BMI < 25
samples = 17
count = [10, 5, 2]
BMI < 24
samples = 13
count = [0, 5, 8]
Well-Groomed=1
samples = 12
count = [9, 3, 0]
Well-Groomed=1
samples =5
count = [1, 2, 2]
samples = 6
count = [6, 0, 0]
class = **attractive**
Hip-to-Waist >.84
samples = 6
count = [3, 3, 0]
Hip-to-Waist >.79
samples = 3
count = [1, 2, 0]
samples = 2
count = [0, 0, 2]
class = **unattractive**
samples = 3
count = [3, 0, 0]
class = **attractive**
samples = 3
count = [0, 3, 0]
class = **average**
samples = 1
count = [1, 0, 0]
class = **attractive**
samples = 2
count = [0, 2, 0]
class = **average**

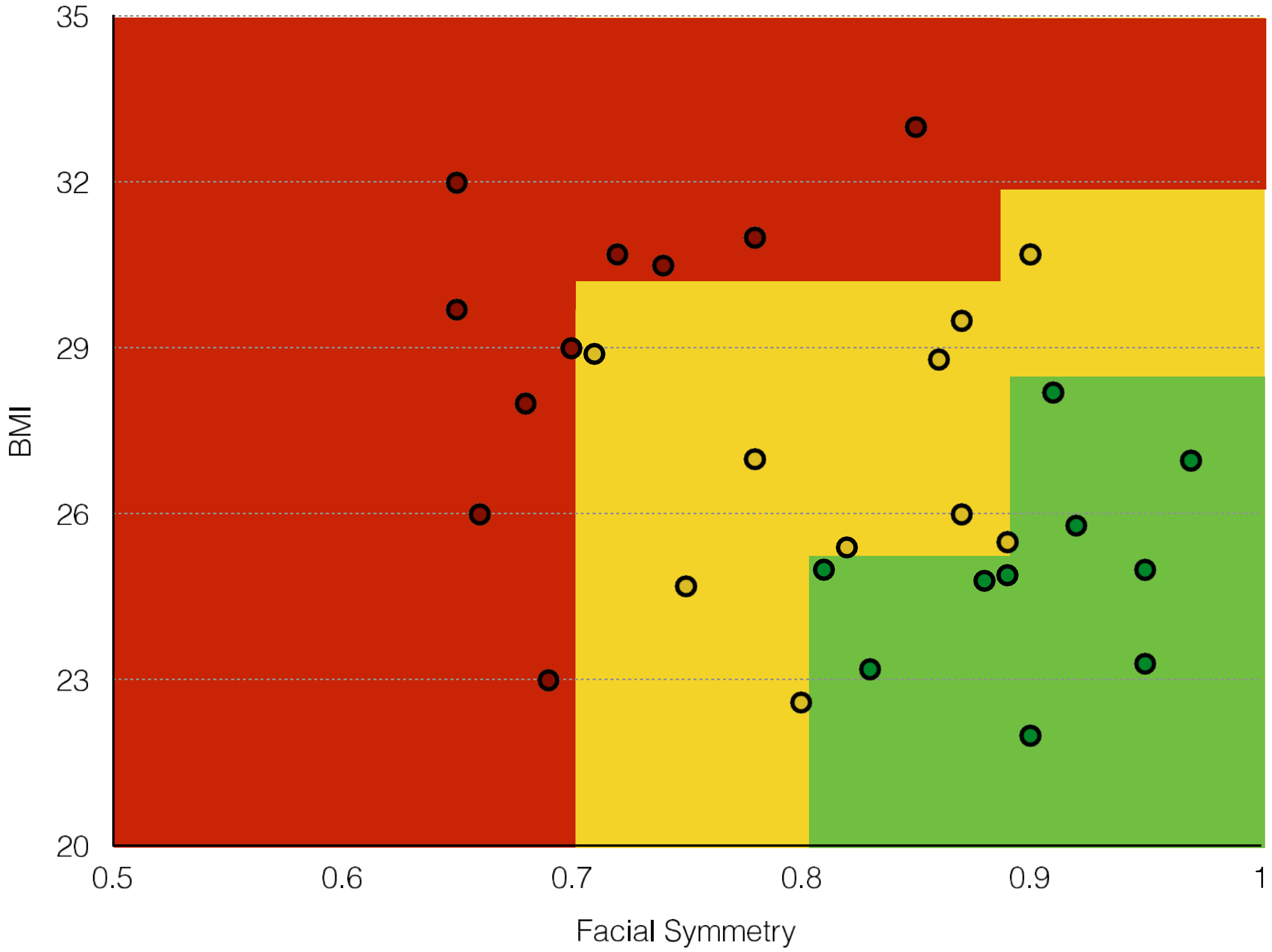
[att, ave, un]


35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry



35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry


35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry



35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry

K-nearest Neighbor


WORK HARD
NAP HARD

	Number of Relationships	Grade of First Relationship
sample a	3	7
sample b	6	11
sample c	3	9
sample d	5	10

Euclidean Distance

point a = [a1, a2]

point b = [b1, b2]

Euclidean Distance

point a = [a1, a2]

point b = [b1, b2]

Two dimensions (features)

$$\sqrt{(a_1-b_1)^2+(a_2-b_2)^2}$$

Euclidean Distance

Feature Vector

$a = [3, 7]$

$b = [6, 11]$

Euclidean Distance

$$\sqrt{(3-6)^2+(7-11)^2}$$

Feature Vector

a = [3, 7]

b = [6, 11]

Euclidean Distance

$$\sqrt{(3-6)^2+(7-11)^2}$$

Feature Vector

a = [3, 7]

b = [6, 11]

$$\sqrt{(-3)^2+(-4)^2}$$

Euclidean Distance

$$\sqrt{(3-6)^2+(7-11)^2}$$

Feature Vector

a = [3, 7]

b = [6, 11]

$$\sqrt{(-3)^2+(-4)^2}$$

$$\sqrt{9+16}$$

Euclidean Distance

$$\sqrt{(3-6)^2+(7-11)^2}$$

Feature Vector

a = [3, 7]

b = [6, 11]

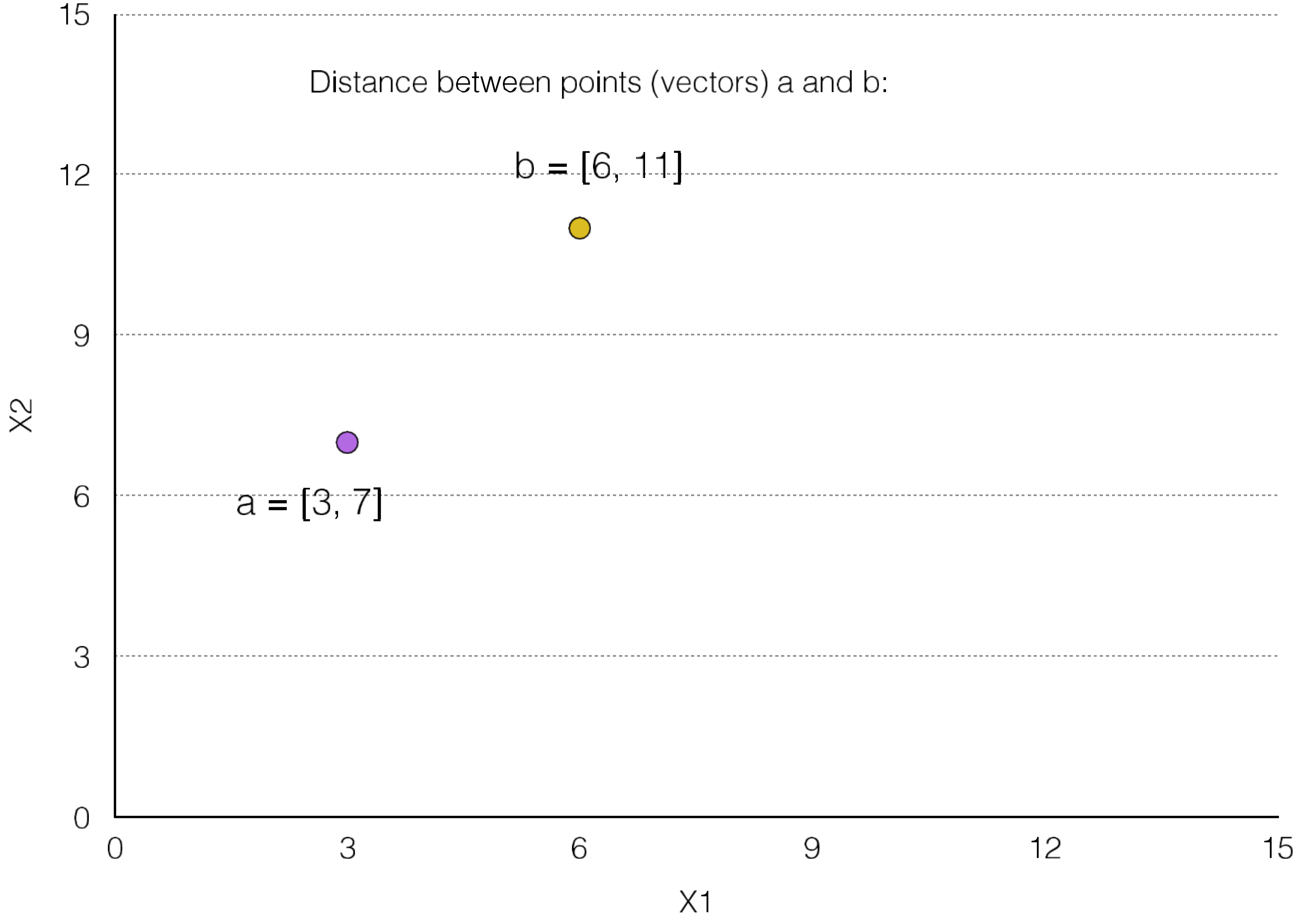
$$\sqrt{(-3)^2+(-4)^2}$$

$$\sqrt{9+16}$$

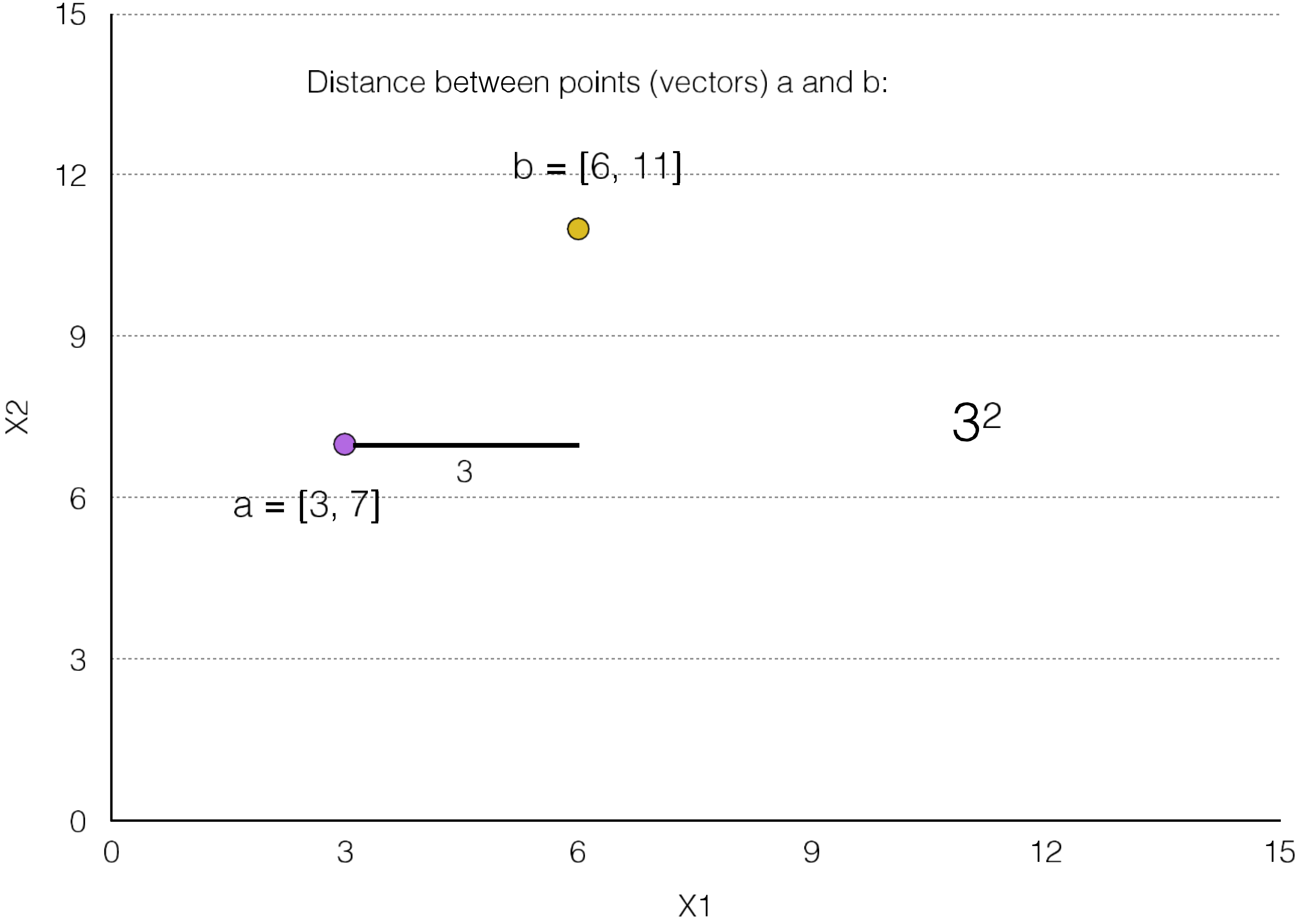
Distance between points (vectors) a and b:

$$\sqrt{25}=5$$

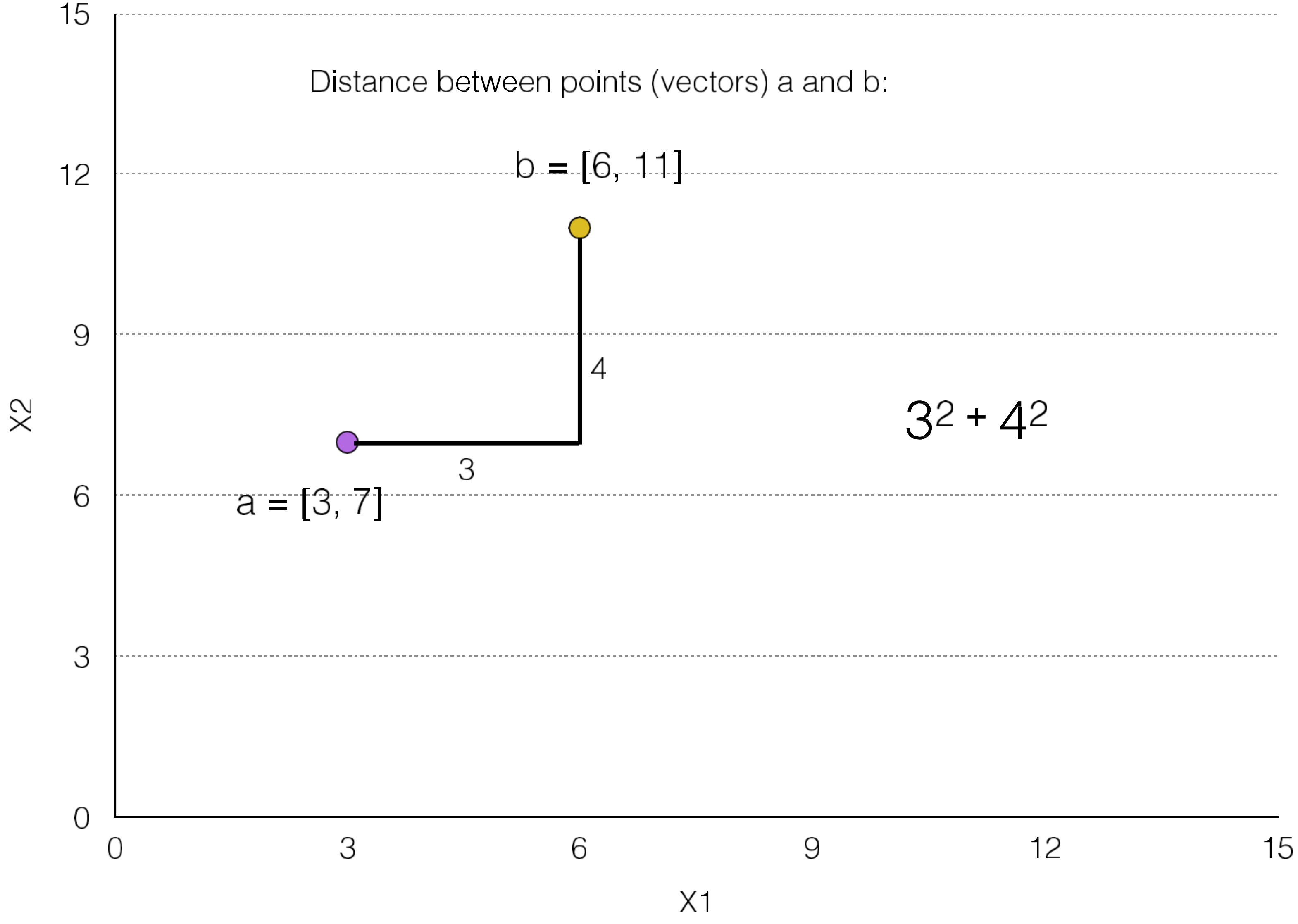
Euclidean Distance



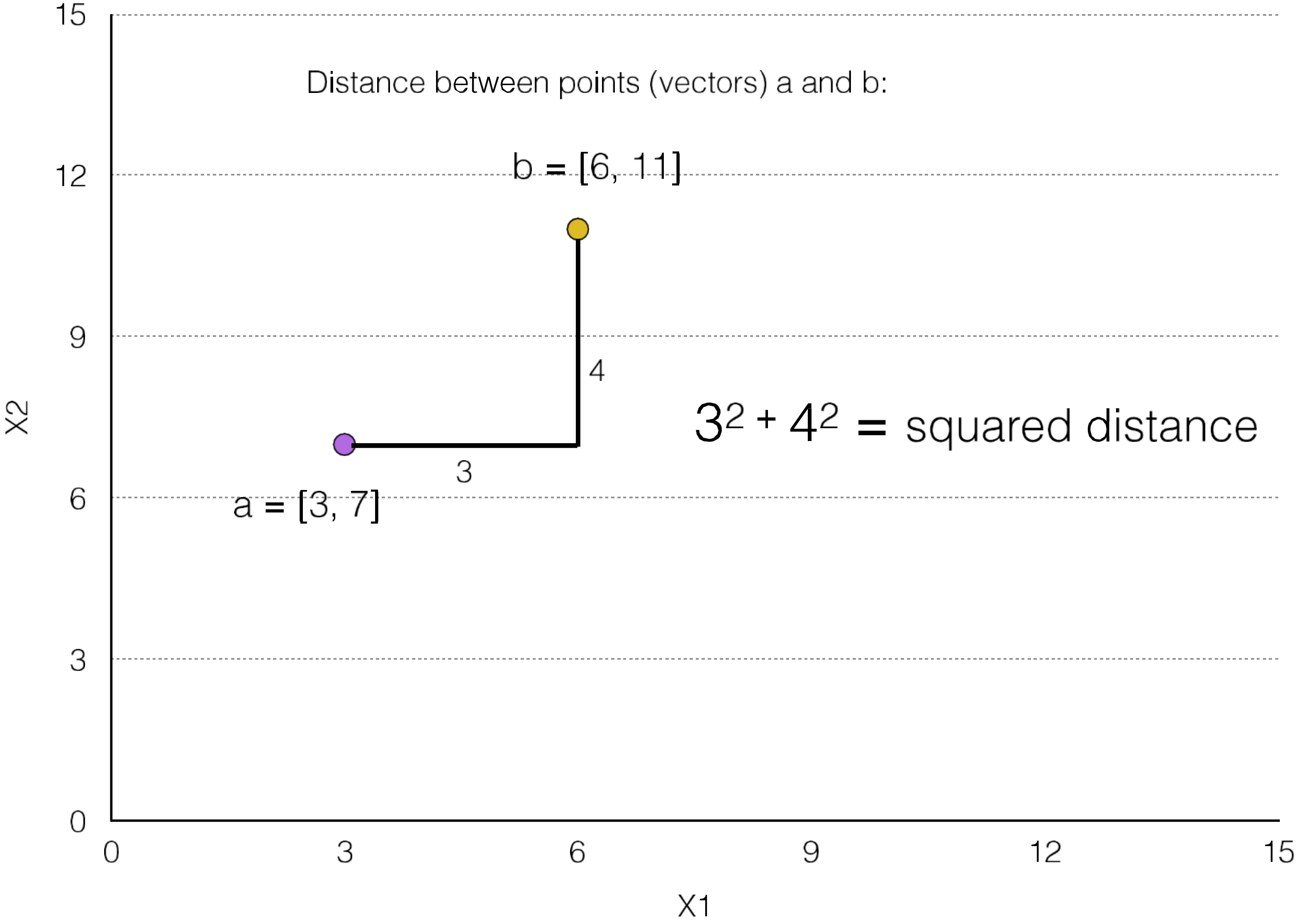
Euclidean Distance



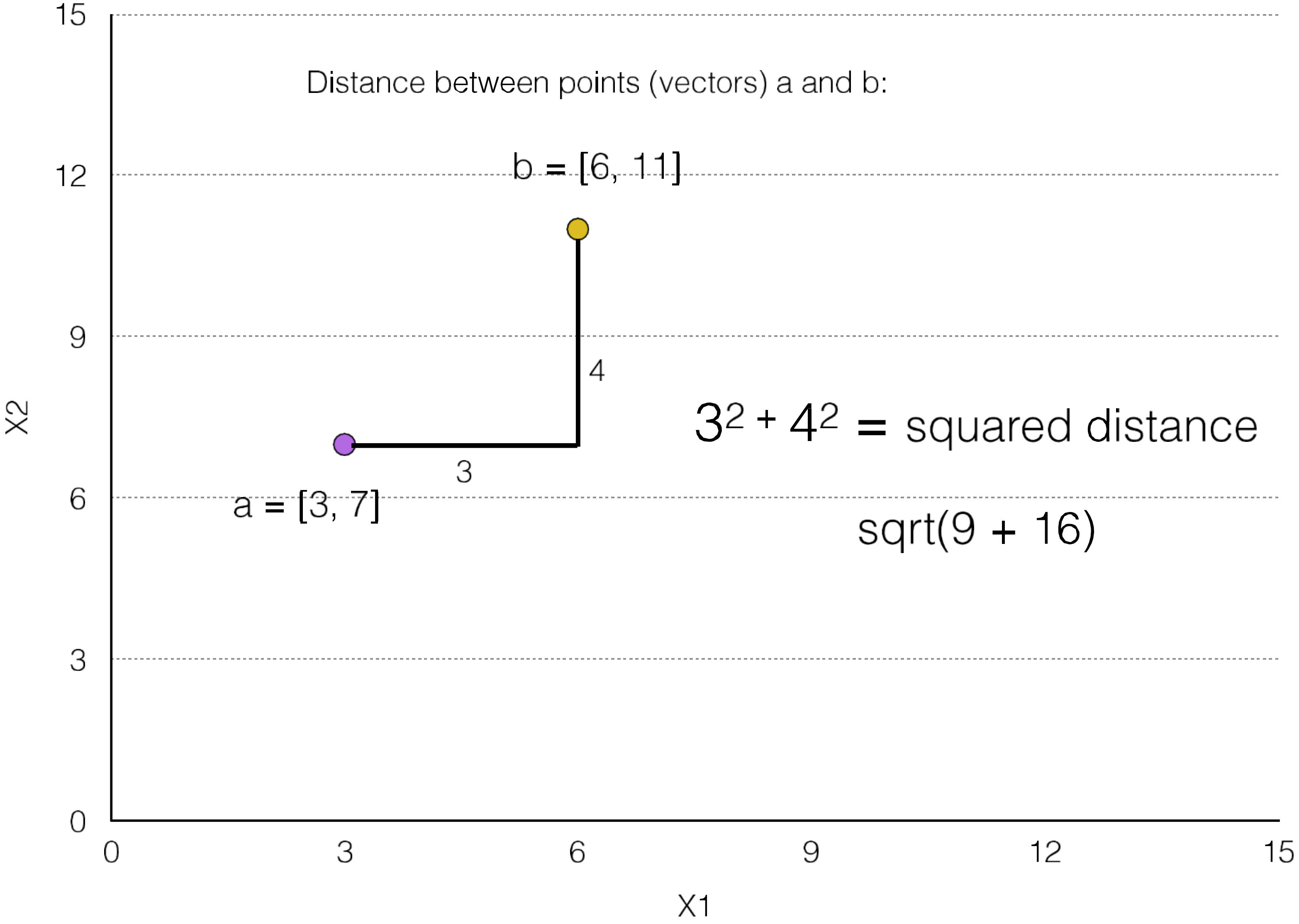
Euclidean Distance



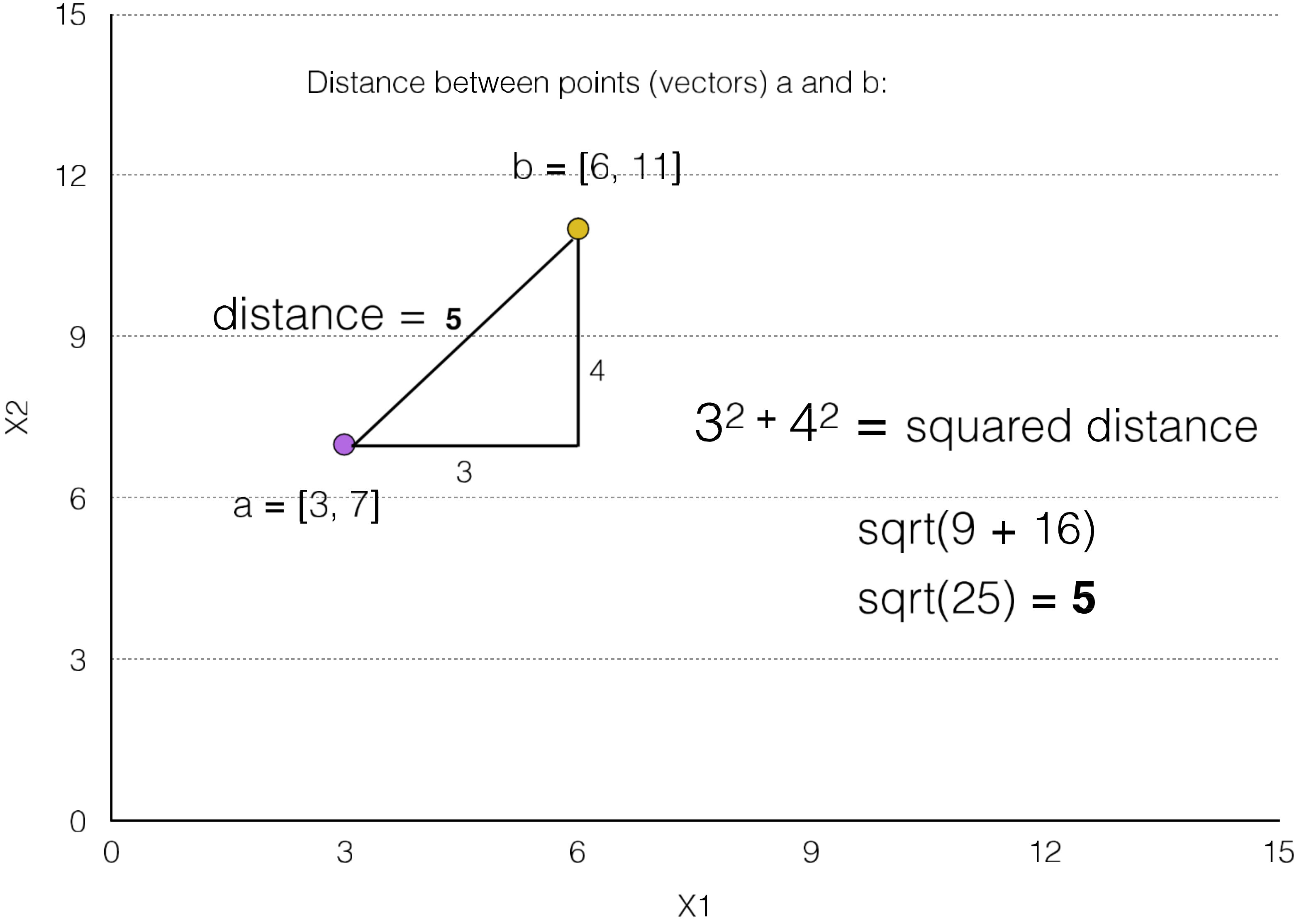
Euclidean Distance



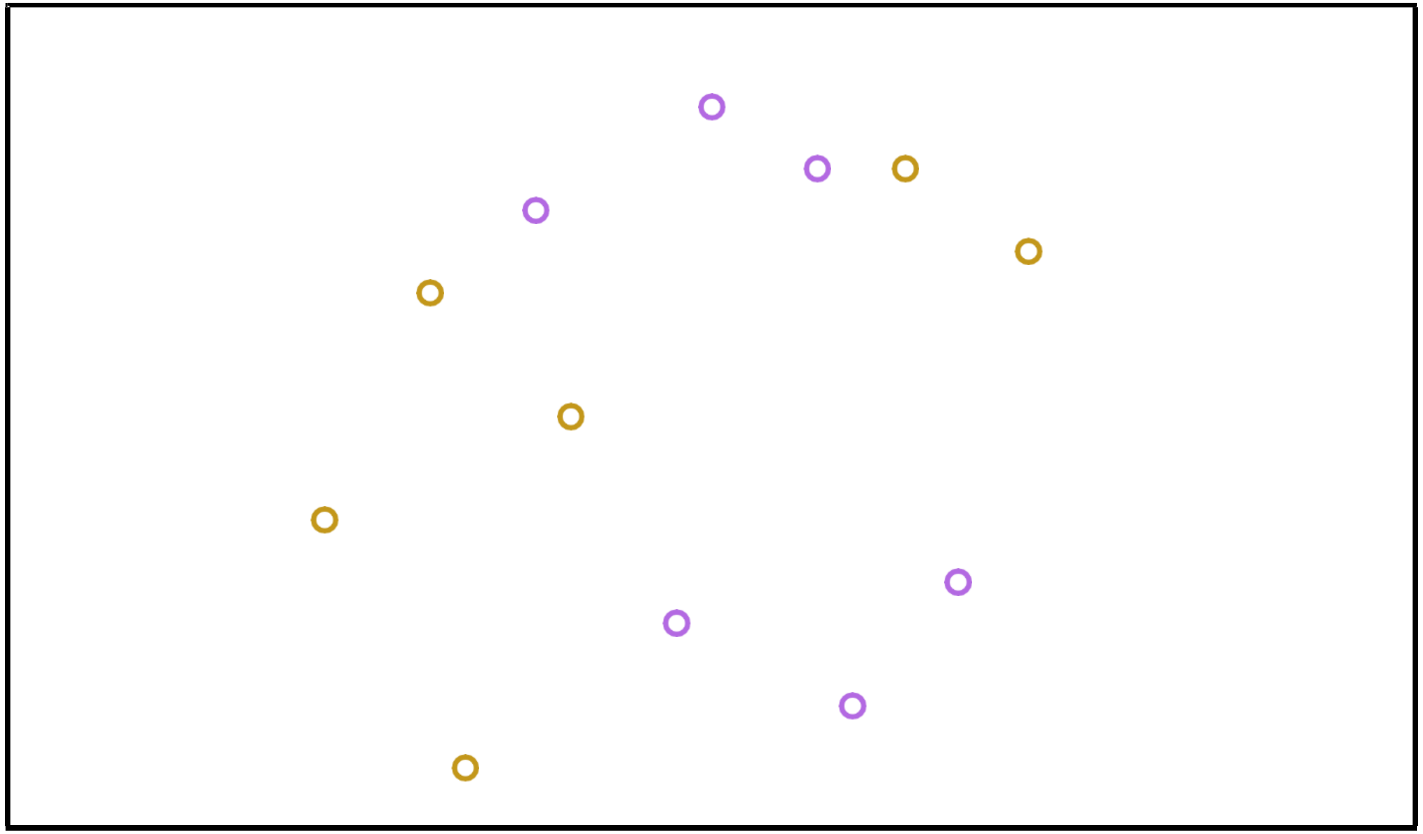
Euclidean Distance



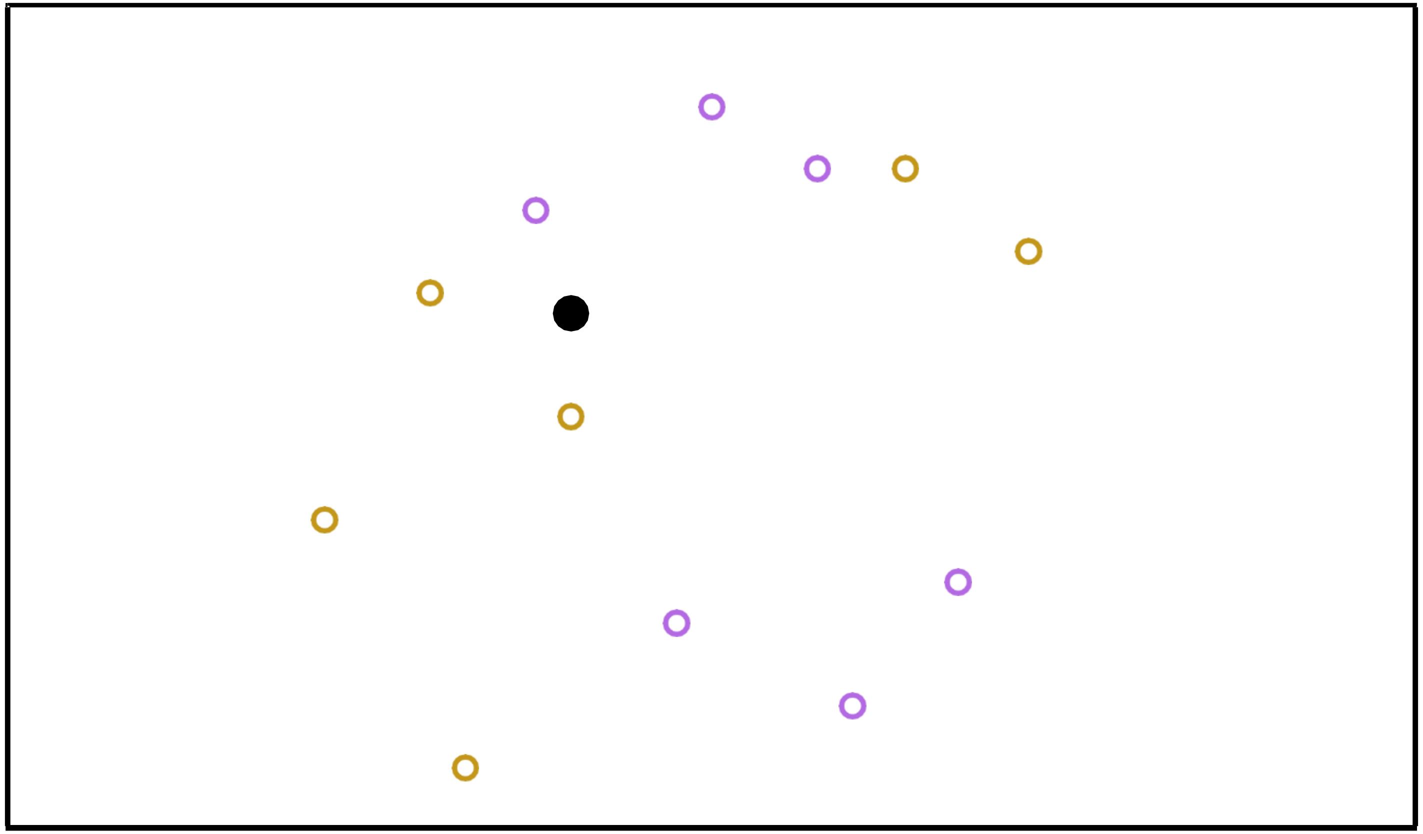
Euclidean Distance



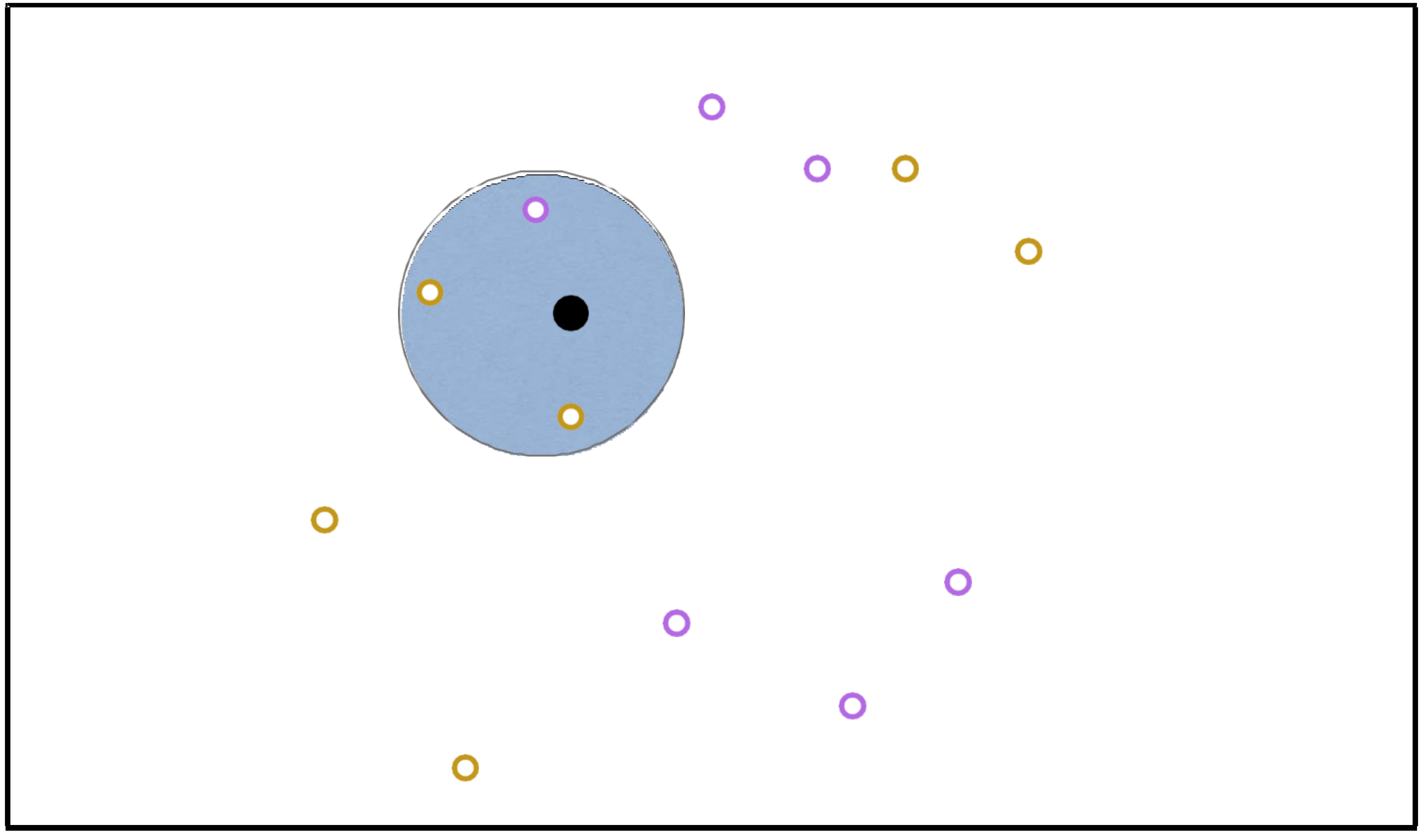
K-nearest Neighbor



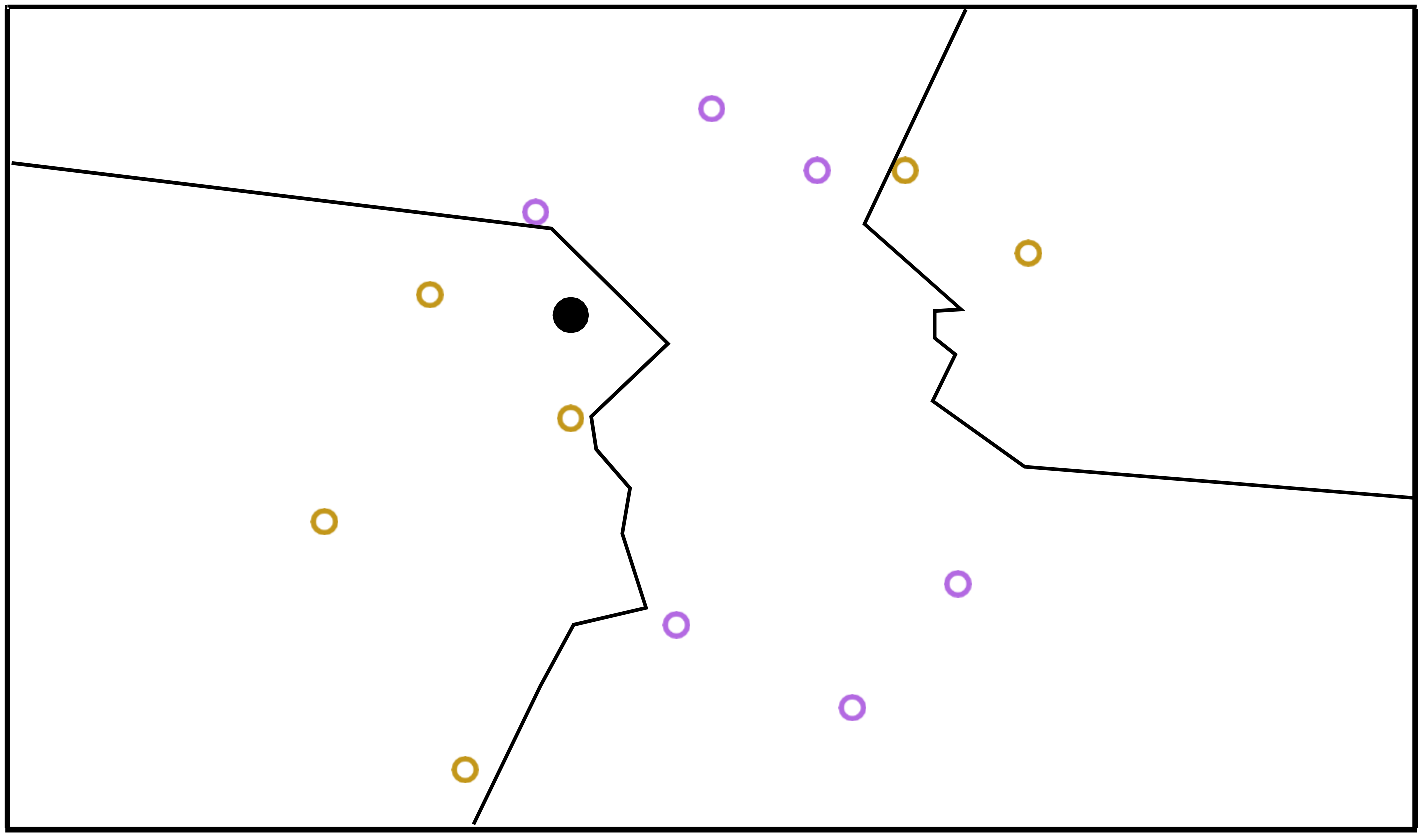
K-nearest Neighbor



K = 3



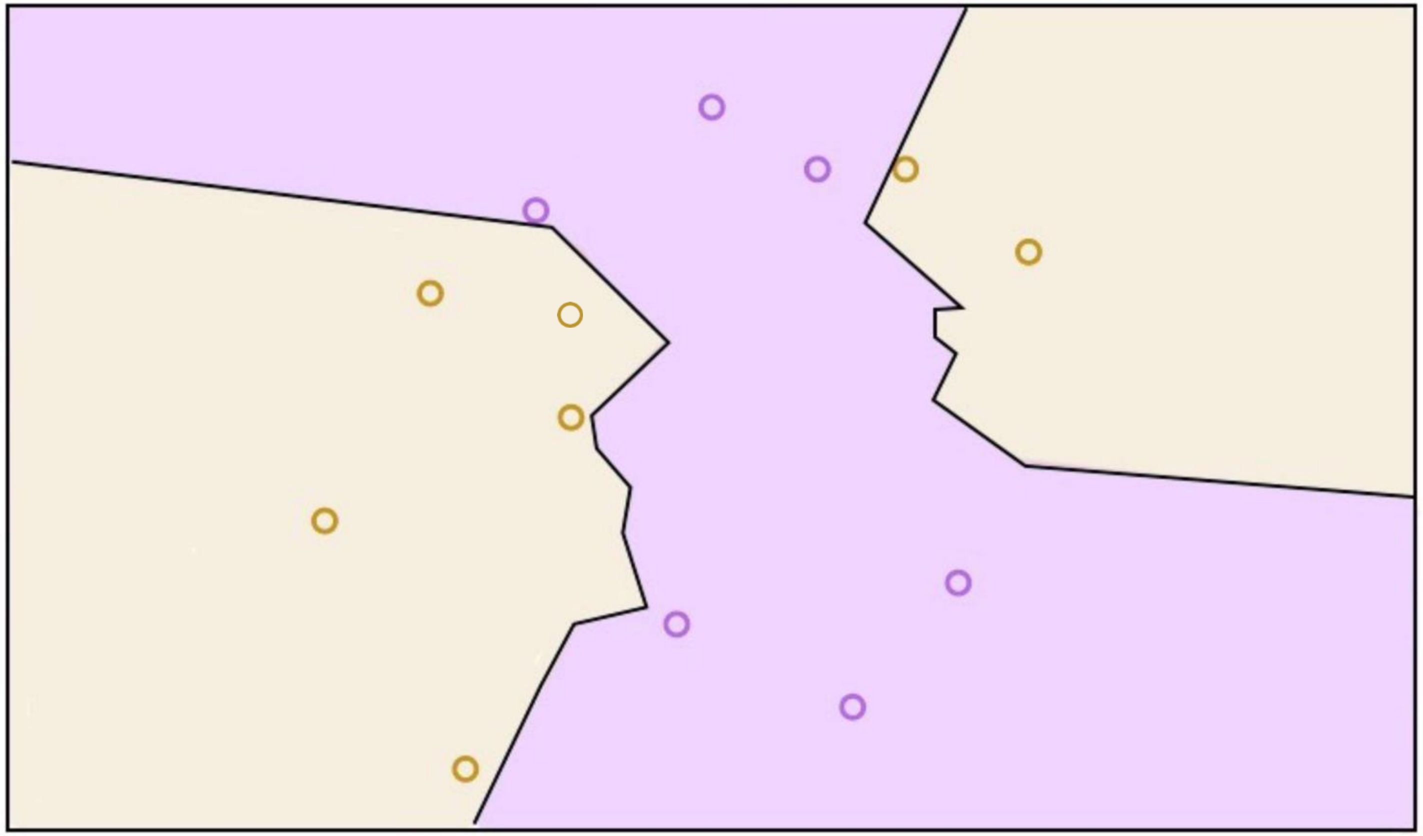
K = 3



Stays together

Breaks up

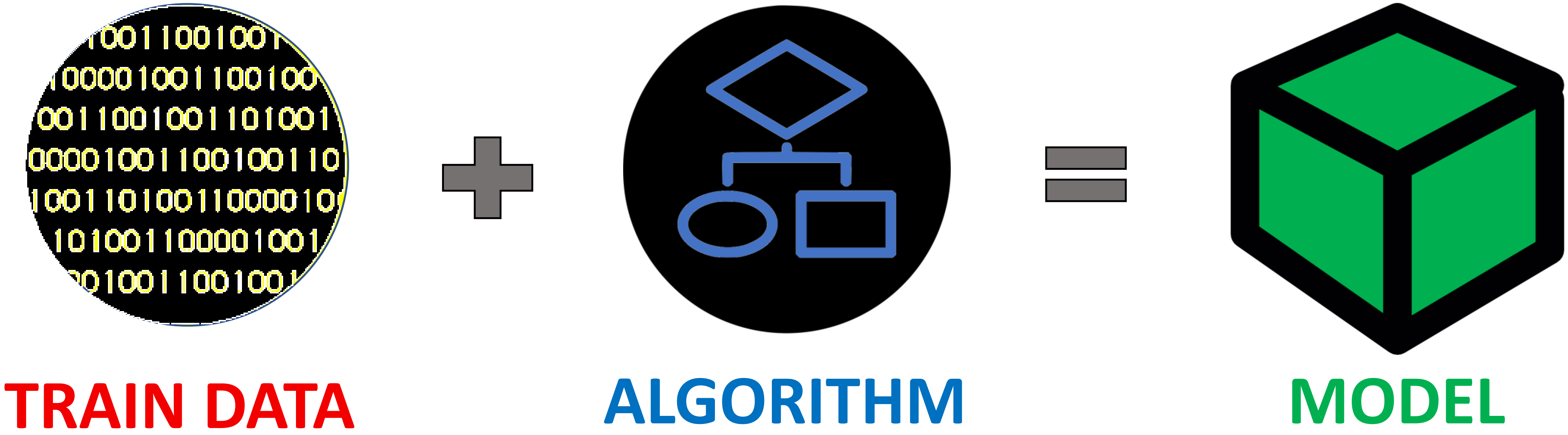
K = 3



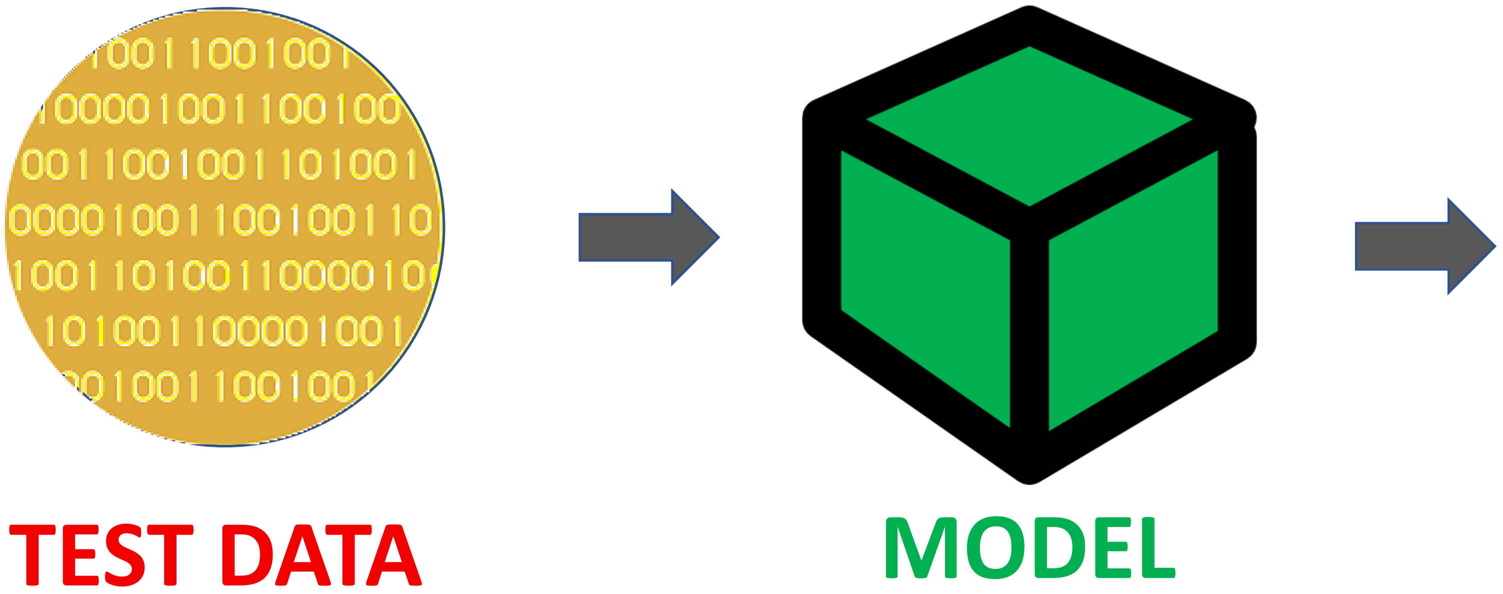
Scikit-learn

```
Model.fit(X_train,y_train)

Model.predict(X_test)

Model.score(X_test,y_test)
```

Model.fit(X_train,y_train)

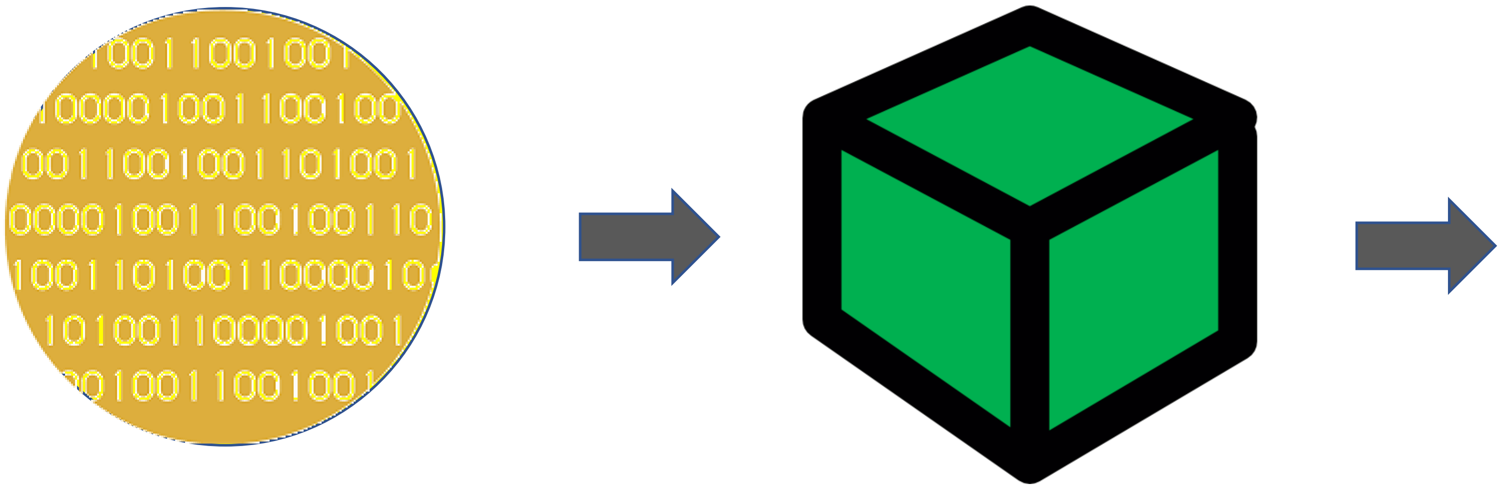


Model. predict(X_test)



Prediction
0
1
0
0
1
0

Model. score(X_test)



TEST DATA

MODEL

Prediction	Label
0	1
1	1
0	0
0	1
1	0
0	0