Acceptance at a University

TEST

GRADES

TEST

GRADES

STUDENT 1
Test: 9/10
Grades: 8/10

STUDENT 2
Test: 3/10
Grades: 4/10

TEST

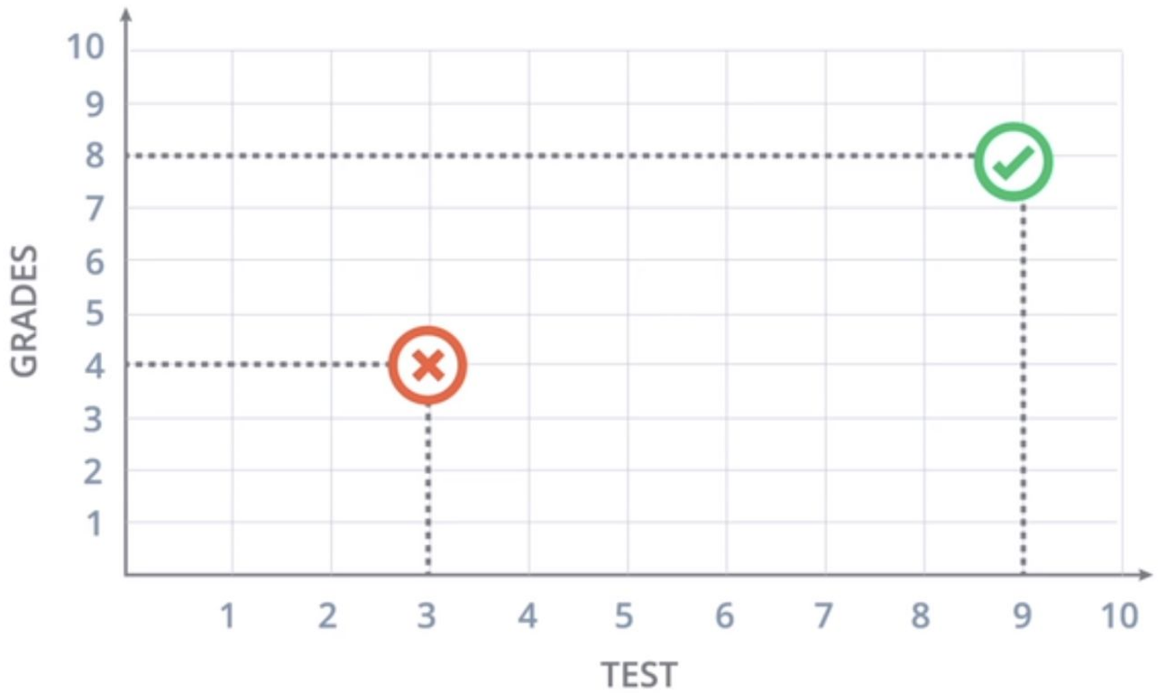
GRADES

STUDENT 1
Test: 9/10
Grades: 8/10

STUDENT 2
Test: 3/10
Grades: 4/10

STUDENT 3
Test: 7/10
Grades: 6/10

STUDENT 3
Test: 7/10
Grades: 6/10



STUDENT 3
Test: 7/10
Grades: 6/10

STUDENT 3
Test: 7/10
Grades: 6/10

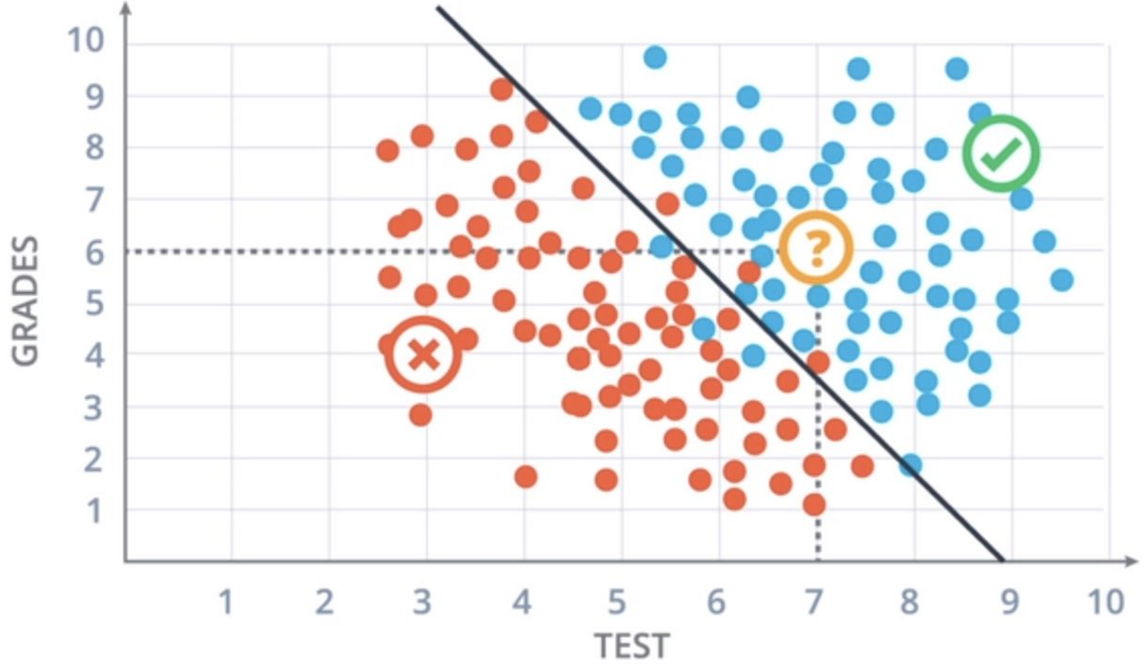
STUDENT 3
Test: 7/10
Grades: 6/10

QUIZ

Does the student get Accepted?

- Yes
- No

Acceptance at a University

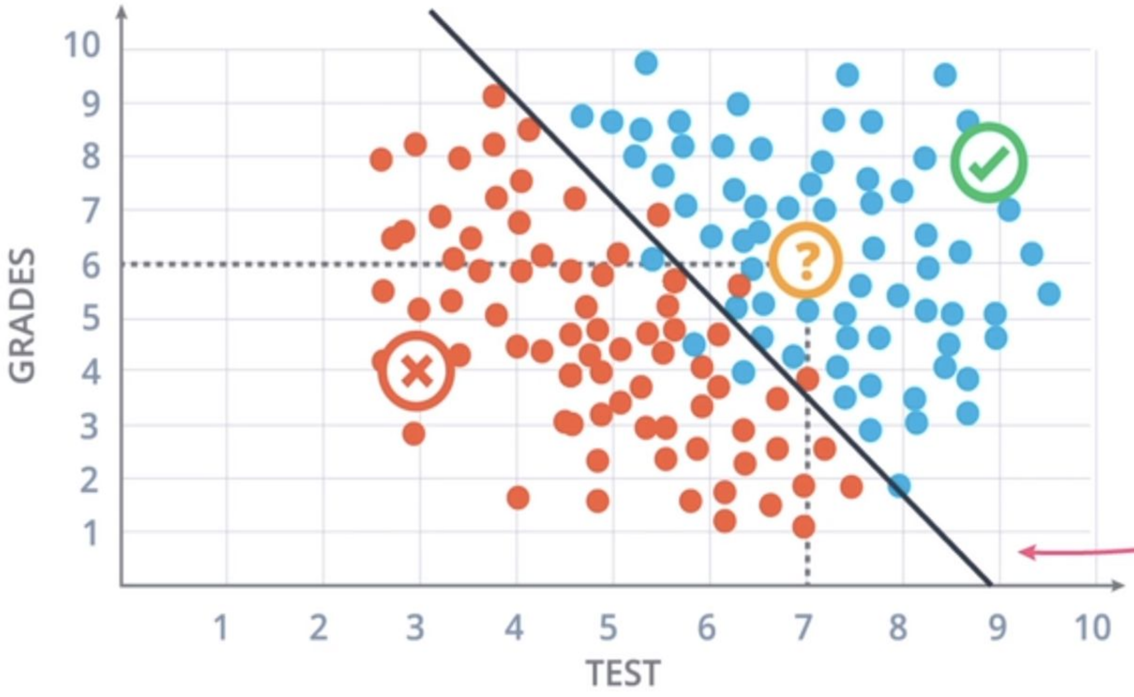


QUIZ

Does the student get Accepted?

- ○ Yes
- ○ No

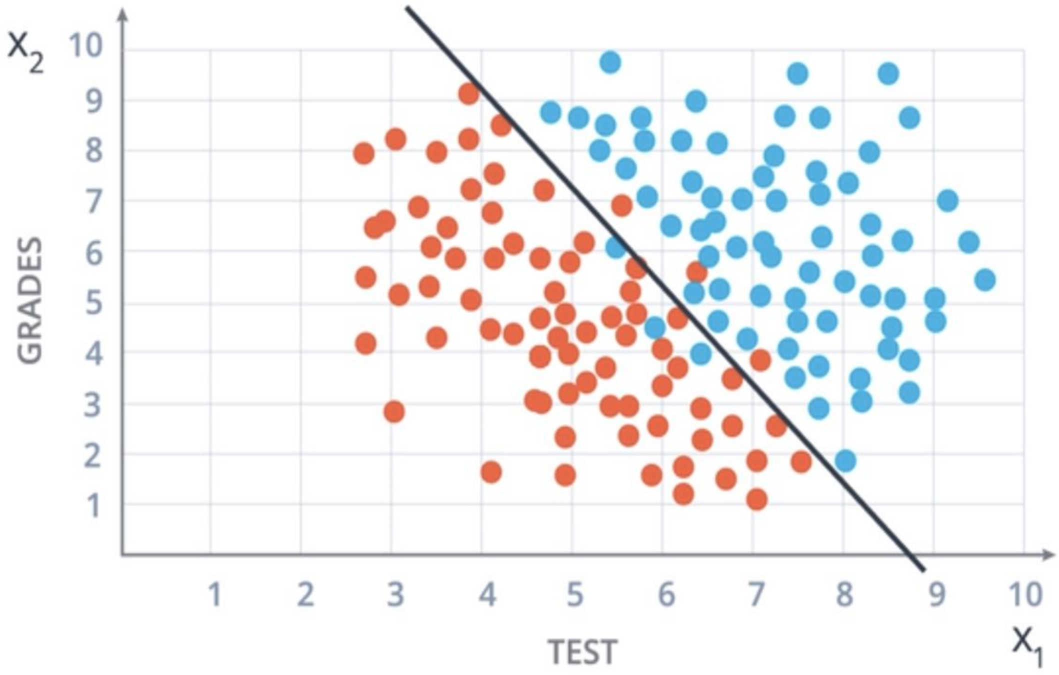
Question



QUESTION

How do we find this line?

Acceptance at a University



BOUNDARY:

A LINE

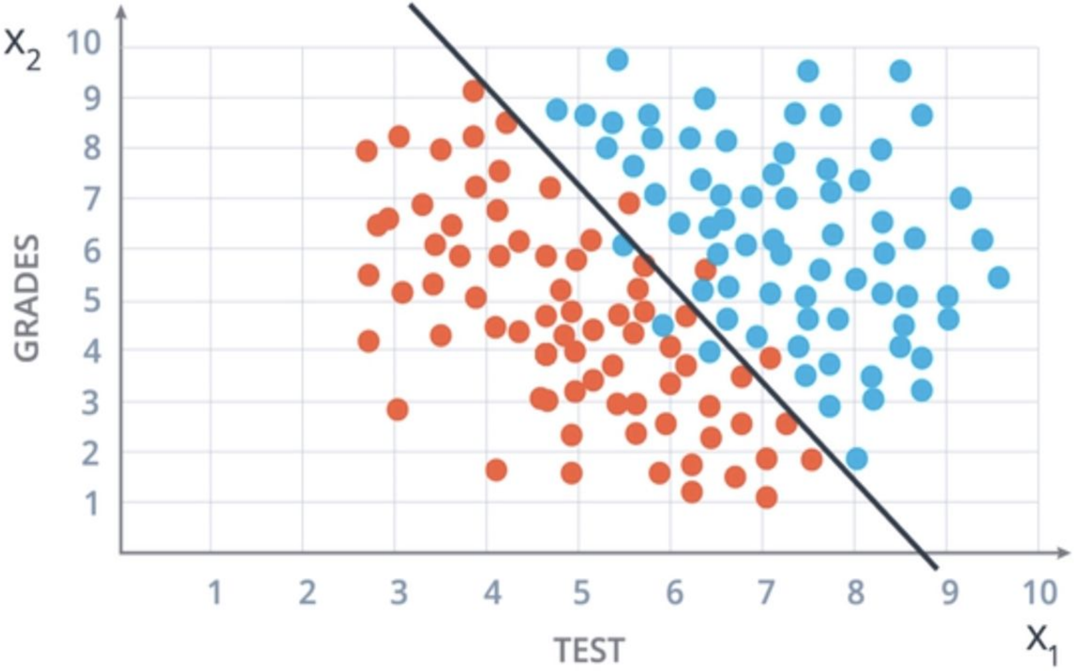
$w_1x_1 + w_2x_2 + b = 0$

$Wx + b = 0$

$W = (w_1, w_2)$

$x = (x_1, x_2)$

Acceptance at a University



BOUNDARY:
A LINE

$w_1x_1 + w_2x_2 + b = 0$

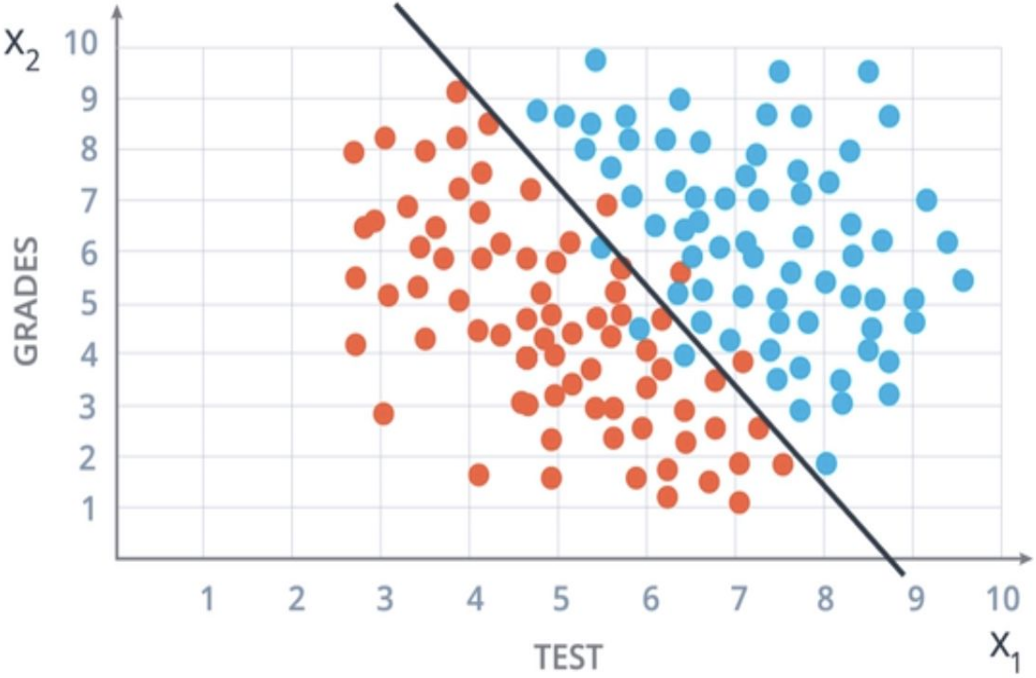
$Wx + b = 0$

$W = (w_1, w_2)$

$x = (x_1, x_2)$

y = label: 0 or 1

Acceptance at a University



BOUNDARY:

A LINE

$w_1x_1 + w_2x_2 + b = 0$

$Wx + b = 0$

$W = (w_1, w_2)$

$x = (x_1, x_2)$

y = label: 0 or 1

PREDICTION:

$$\hat{y} = \begin{cases} 1 \text{ if } Wx + b \geq 0 \\ 0 \text{ if } Wx + b < 0 \end{cases}$$

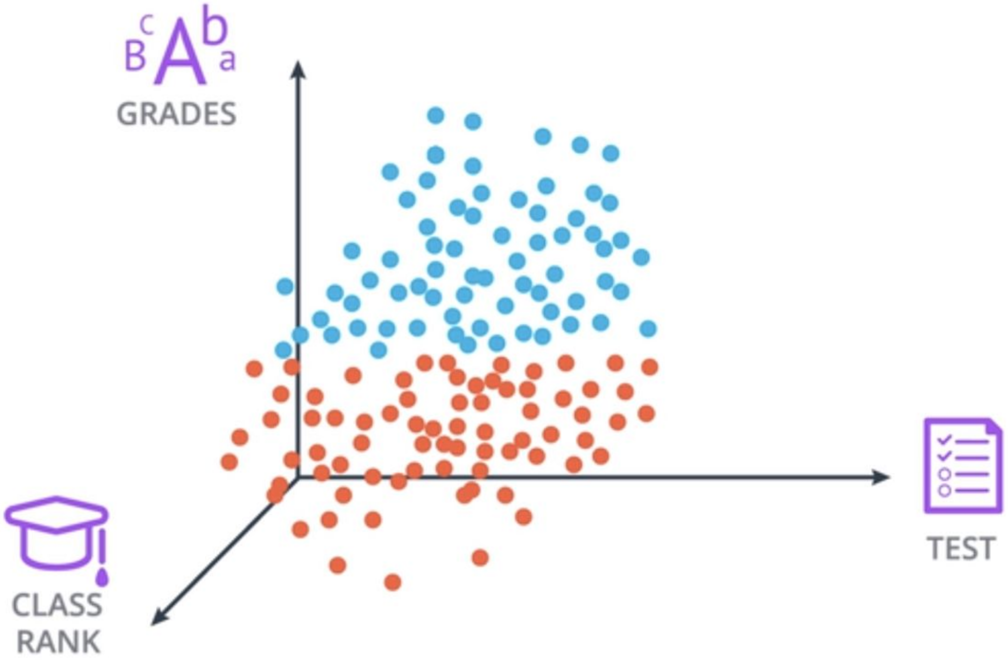
Acceptance at a University

GRADES

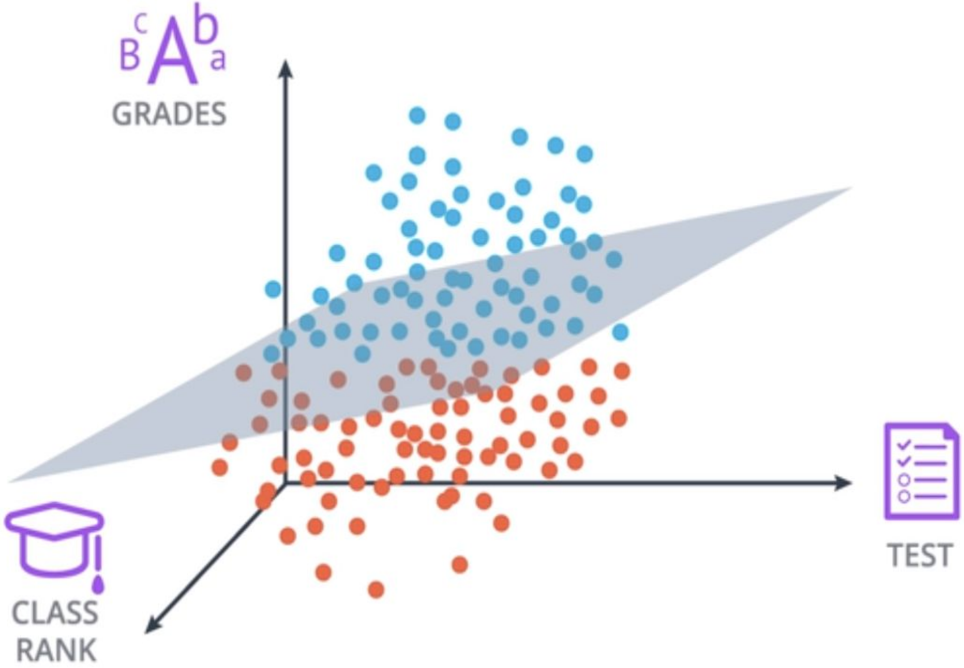
TEST

CLASS RANK

Acceptance at a University



Acceptance at a University



BOUNDARY:

A PLANE

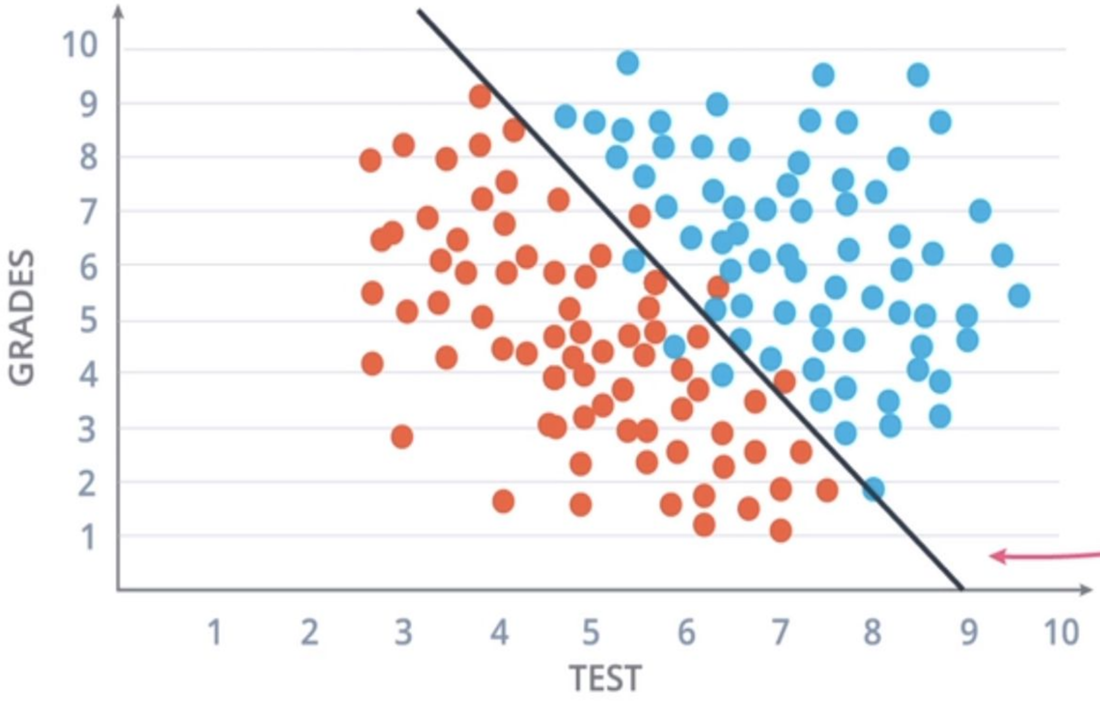
$w_1x_1 + w_2x_2 + w_3x_3 + b = 0$

$Wx + b = 0$

PREDICTION:

$$\hat{y} = \begin{cases} 1 \text{ if } Wx + b \geq 0 \\ 0 \text{ if } Wx + b < 0 \end{cases}$$

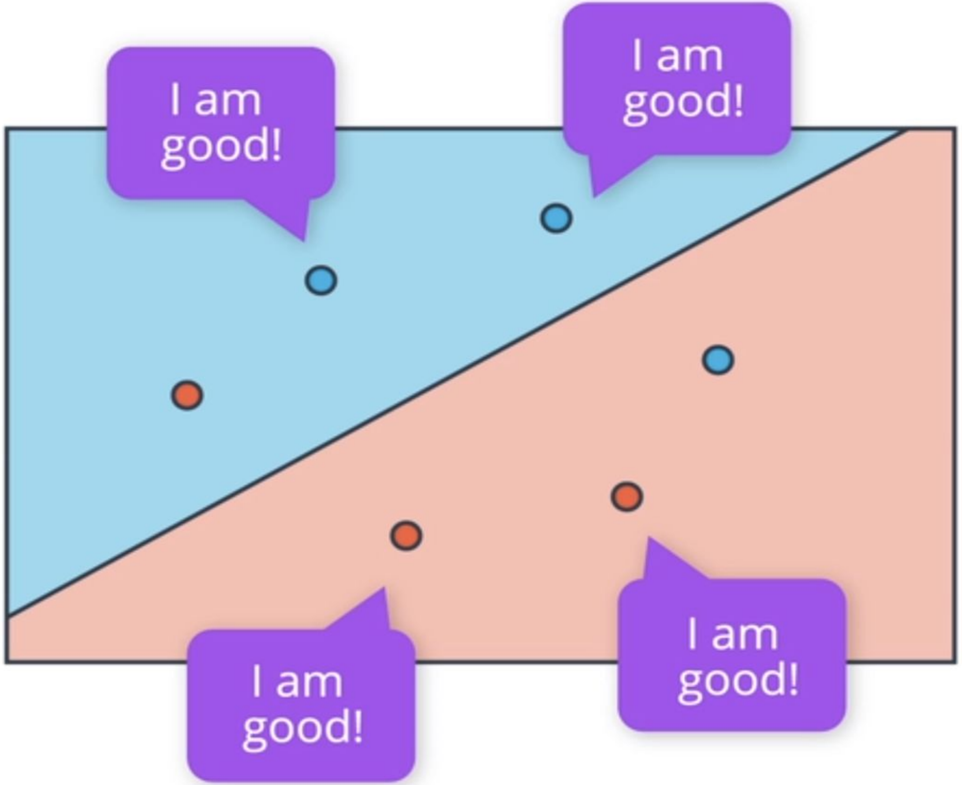
Question



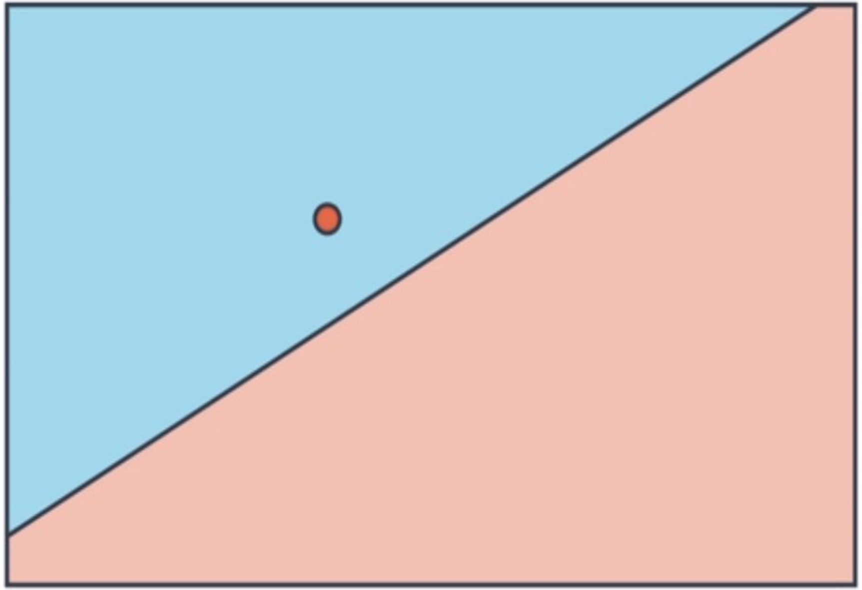
QUESTION

How do we find this line?

Goal: Split Data



Goal: Split Data

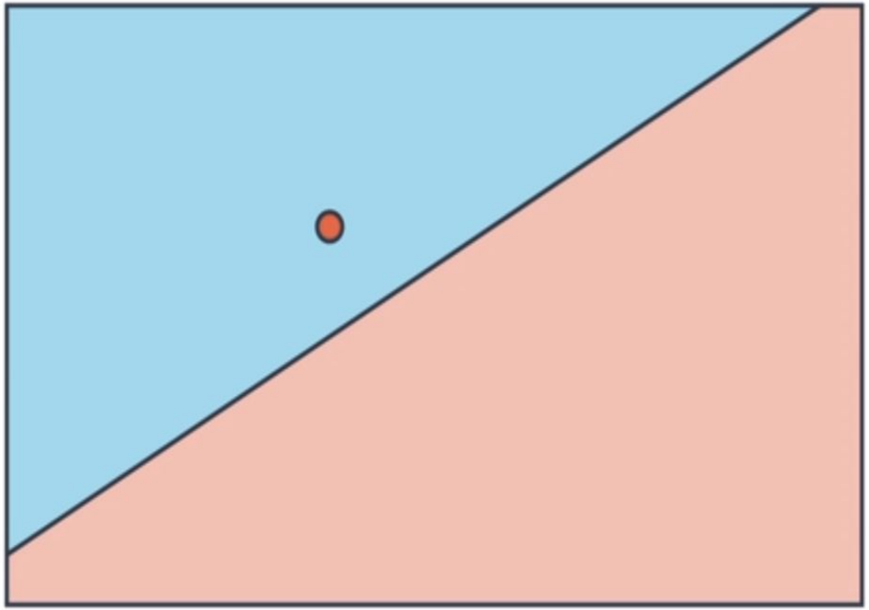


QUIZ

Does the misclassified point want the line to be close or farther?

- ○ Closer
- ○ Farther

Goal: Split Data

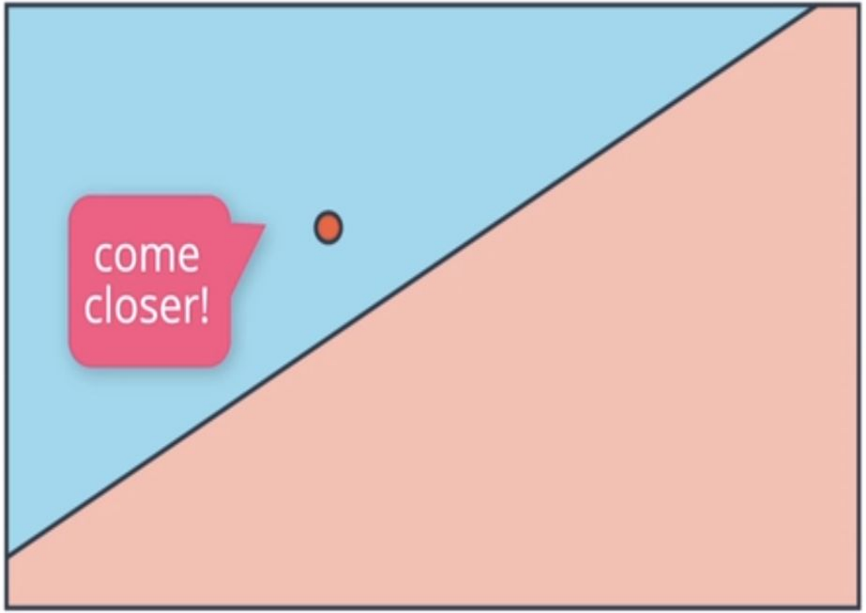


QUIZ

Does the misclassified point want the line to be close or farther?

- Closer
- Farther

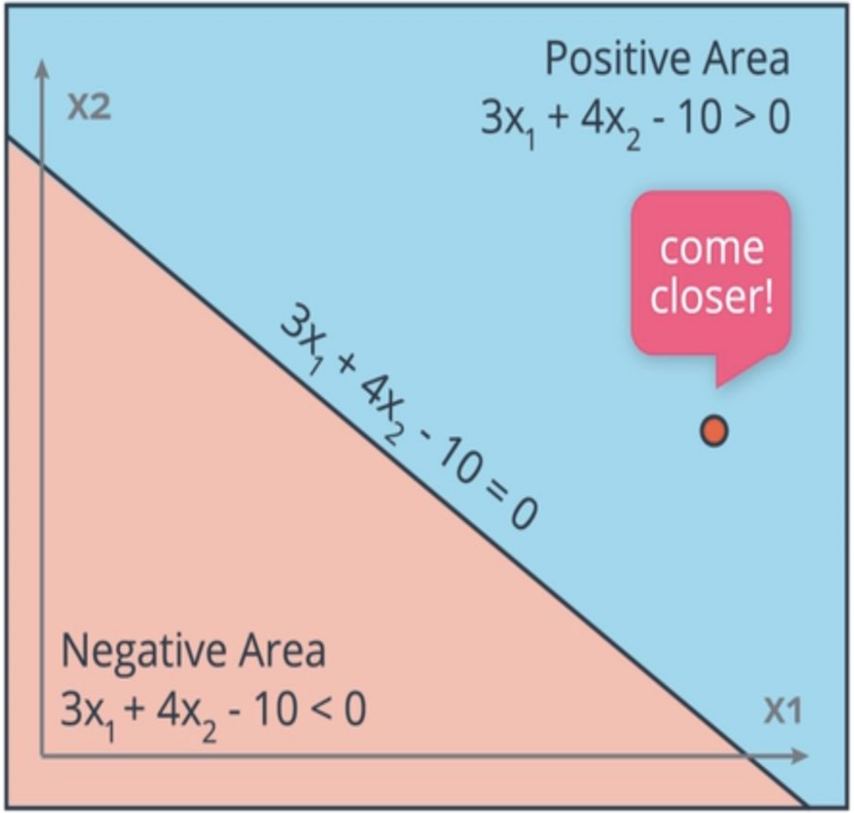
Goal: Split Data



QUIZ

Does the misclassified point want the line to be close or farther?

- Closer
- Farther

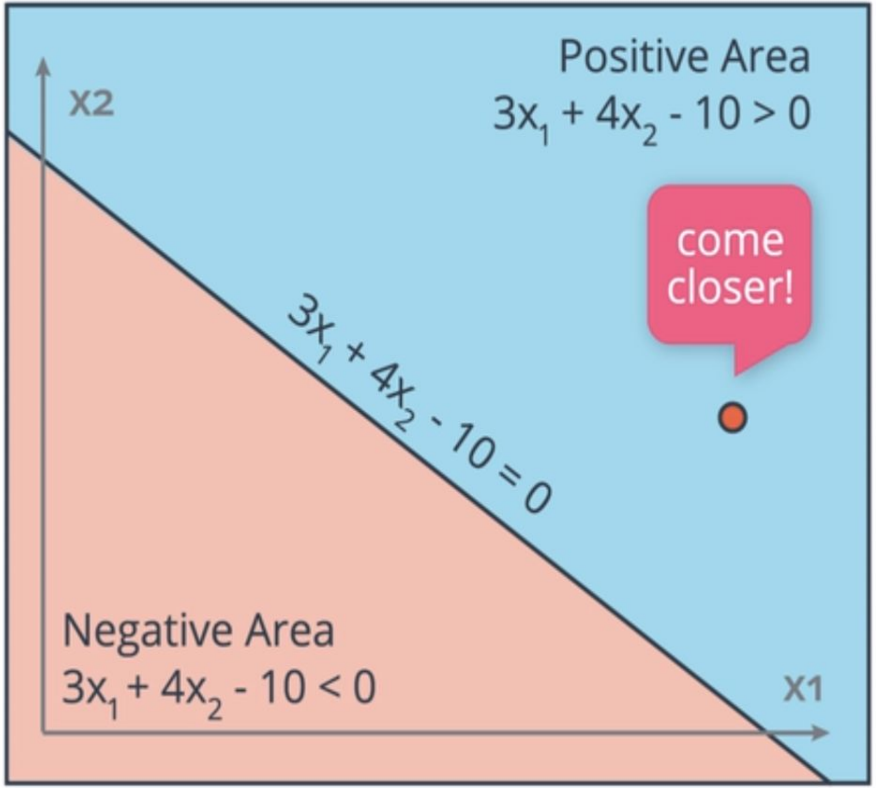


LINE: $3x_1 + 4x_2 - 10 = 0$

POINT: (4,5)

LEARNING RATE: 0.1

	3	4	-10
−	0.4	0.5	0.1
	2.6	3.5	-10.1

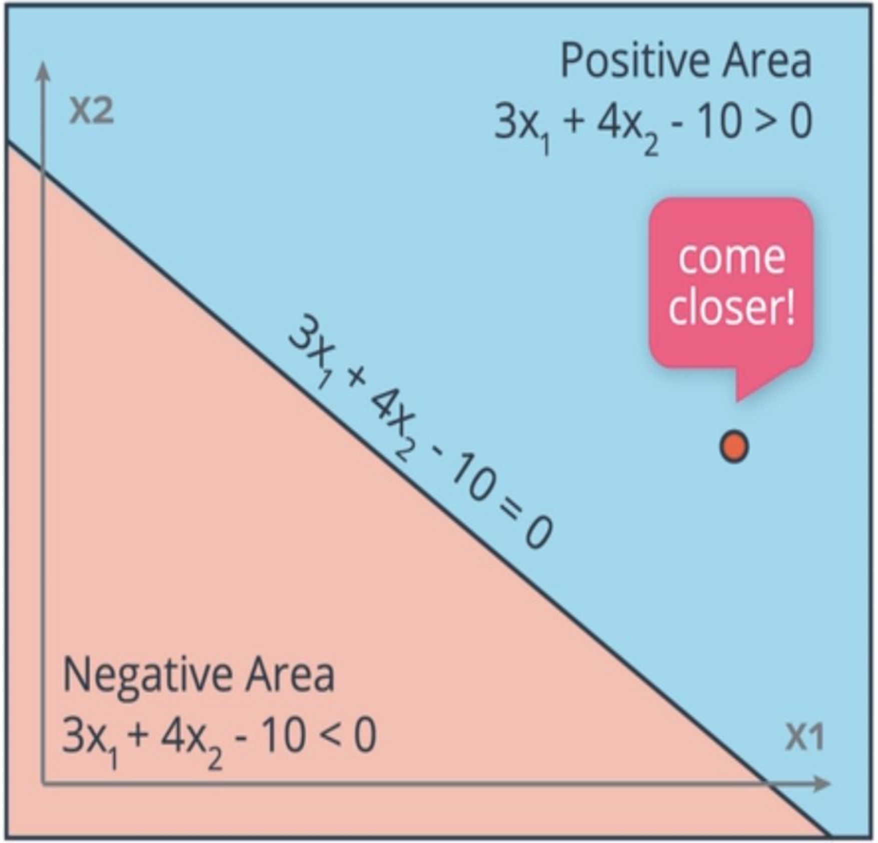


LINE: $3x_1 + 4x_2 - 10 = 0$

POINT: (4,5)

LEARNING RATE: 0.1

	3	4	-10
−	0.4	0.5	0.1
	2.6	3.5	-10.1



LINE: $3x_1 + 4x_2 - 10 = 0$

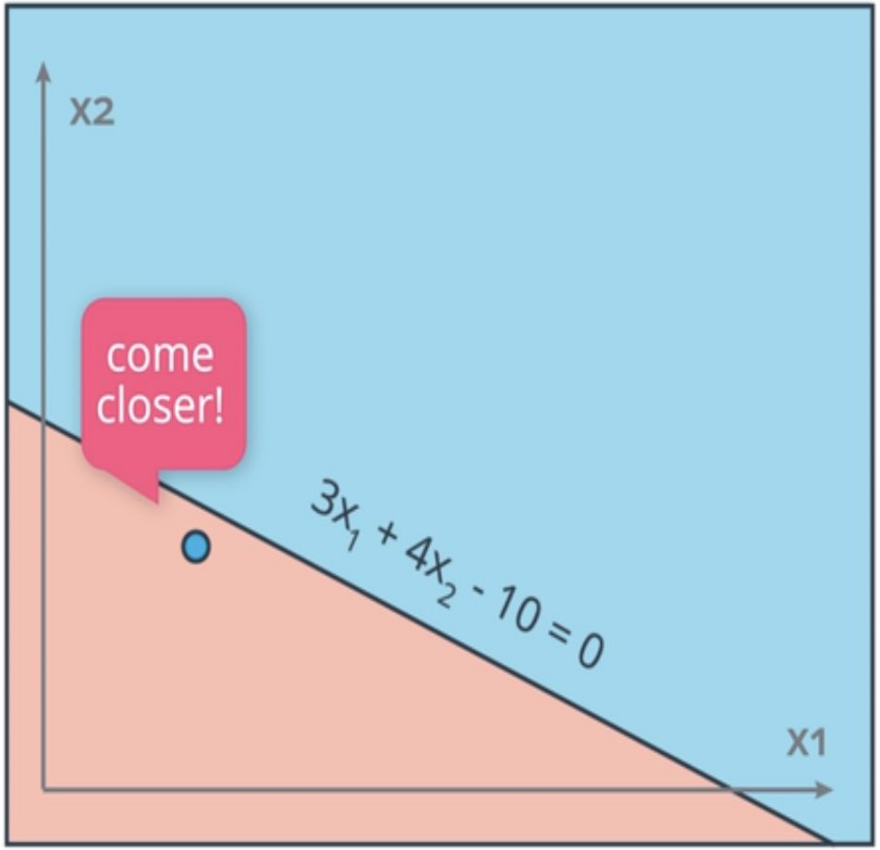
POINT: (4,5)

LEARNING RATE: 0.1

$$
\begin{array}{rccc}
 & 3 & 4 & -10 \\
- & 0.4 & 0.5 & 0.1 \\
\hline
 & 2.6 & 3.5 & -10.1
\end{array}
$$

NEW LINE

$2.6x_1 + 3.5x_2 - 10.1 = 0$



LINE: $3x_1 + 4x_2 - 10 = 0$

POINT: (1,1)

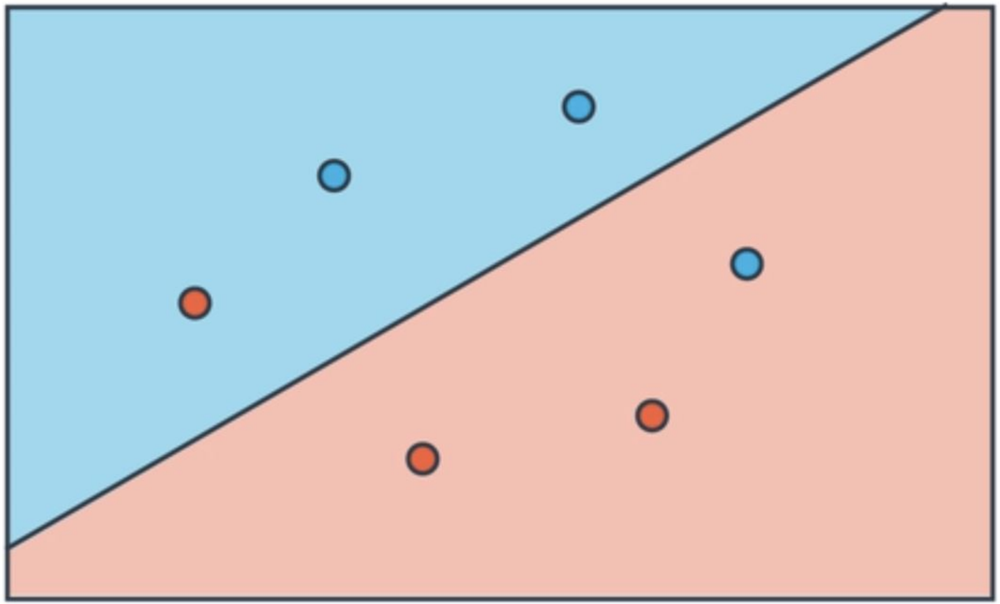
LEARNING RATE: 0.1

$$
\begin{array}{cccc}
 & 3 & 4 & -10 \\
+ & 0.1 & 0.1 & 0.1 \\
\hline
 & 3.1 & 4.1 & -9.9
\end{array}
$$

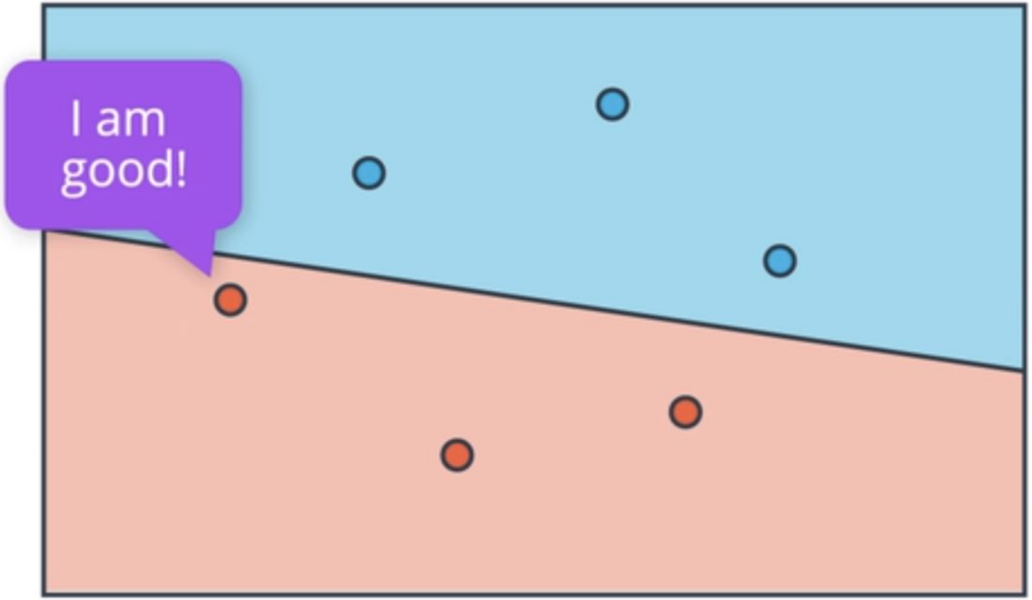
NEW LINE

$3.1x_1 + 4.1x_2 - 9.9 = 0$

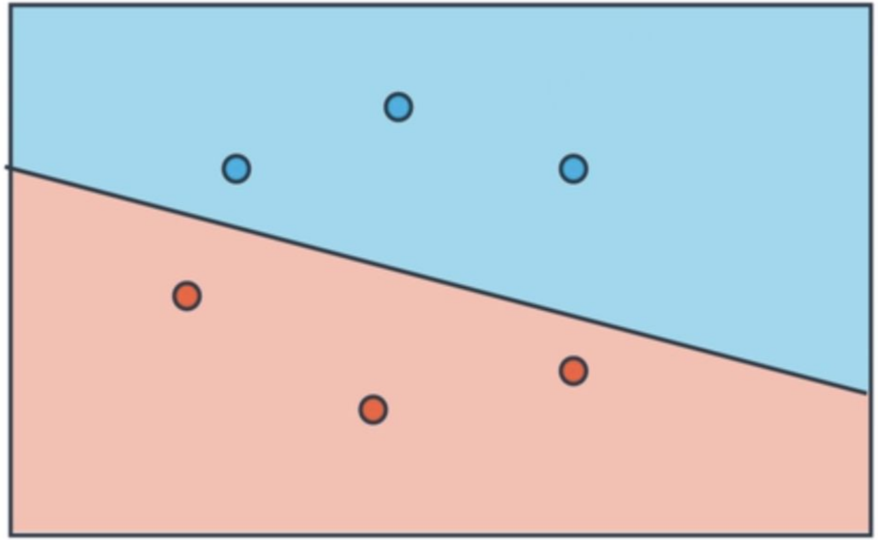
Goal: Split Data



Goal: Split Data



Perceptron Algorithm



1. Start with random weights: $w_1, \ldots, w_n, b$

2. For every misclassified point $(x_1,\ldots,x_n)$:

   2.1. If prediction = 0:
   - For i = 1 …n
     - Change $w_i + \alpha x_i$
   - Change b to $b + \alpha$

   2.2. If prediction = 1:
   - For i = 1 …n
     - Change $w_i - \alpha x_i$
   - Change b to $b - \alpha$

Acceptance at a University

	x_1	x_2	x_3		x_n	y
	EXAM 1	EXAM 2	GRADES	...	ESSAY	PASS?
STUDENT 1	9	6	5	...	6	1(yes)
STUDENT 2	8	4	8	...	3	0(no)
...	...	...	...	...	...	
STUDENT n	6	7	2	...	8	1(yes)

← n columns →

n-dimensional space

$x_1, x_2, \ldots, x_n$

BOUNDARY:

n-1 dimensional hyperplane

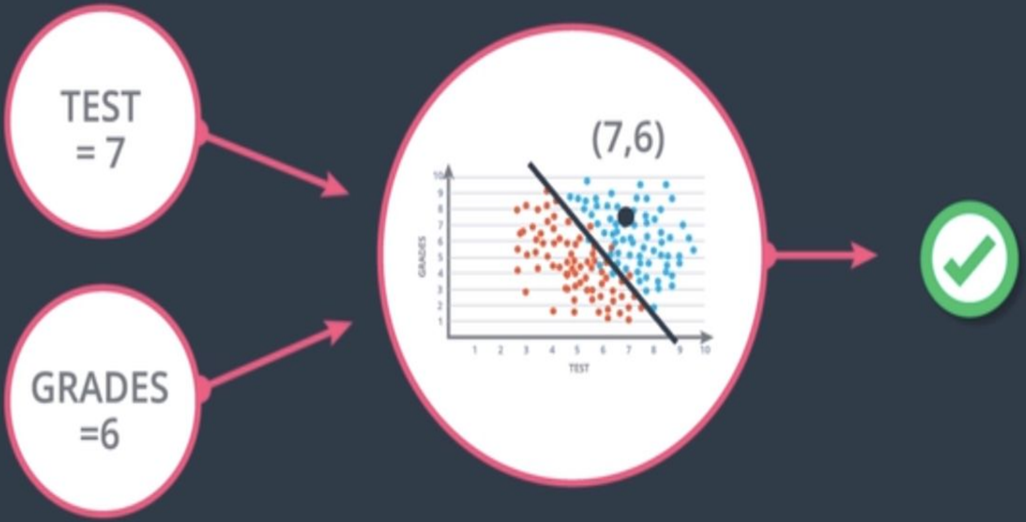
$w_1x_1 + w_2x_2 + w_nx_n + b = 0$

$Wx + b = 0$

PREDICTION:

$$\hat{y} = \begin{cases} 1 \text{ if } Wx + b \geq 0 \\ 0 \text{ if } Wx + b < 0 \end{cases}$$

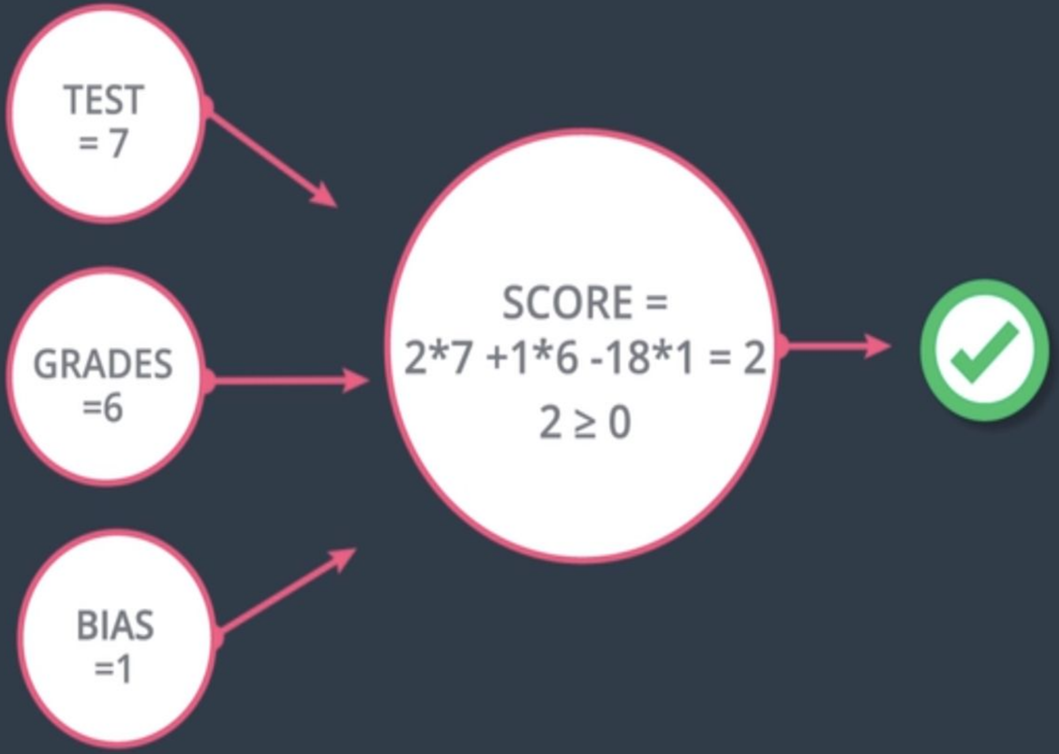
Perceptron



Score=
2*Test + 1*Grades -18

PREDICTION:
Score ≥ 0 Accept
Score < 0 Reject

Perceptron



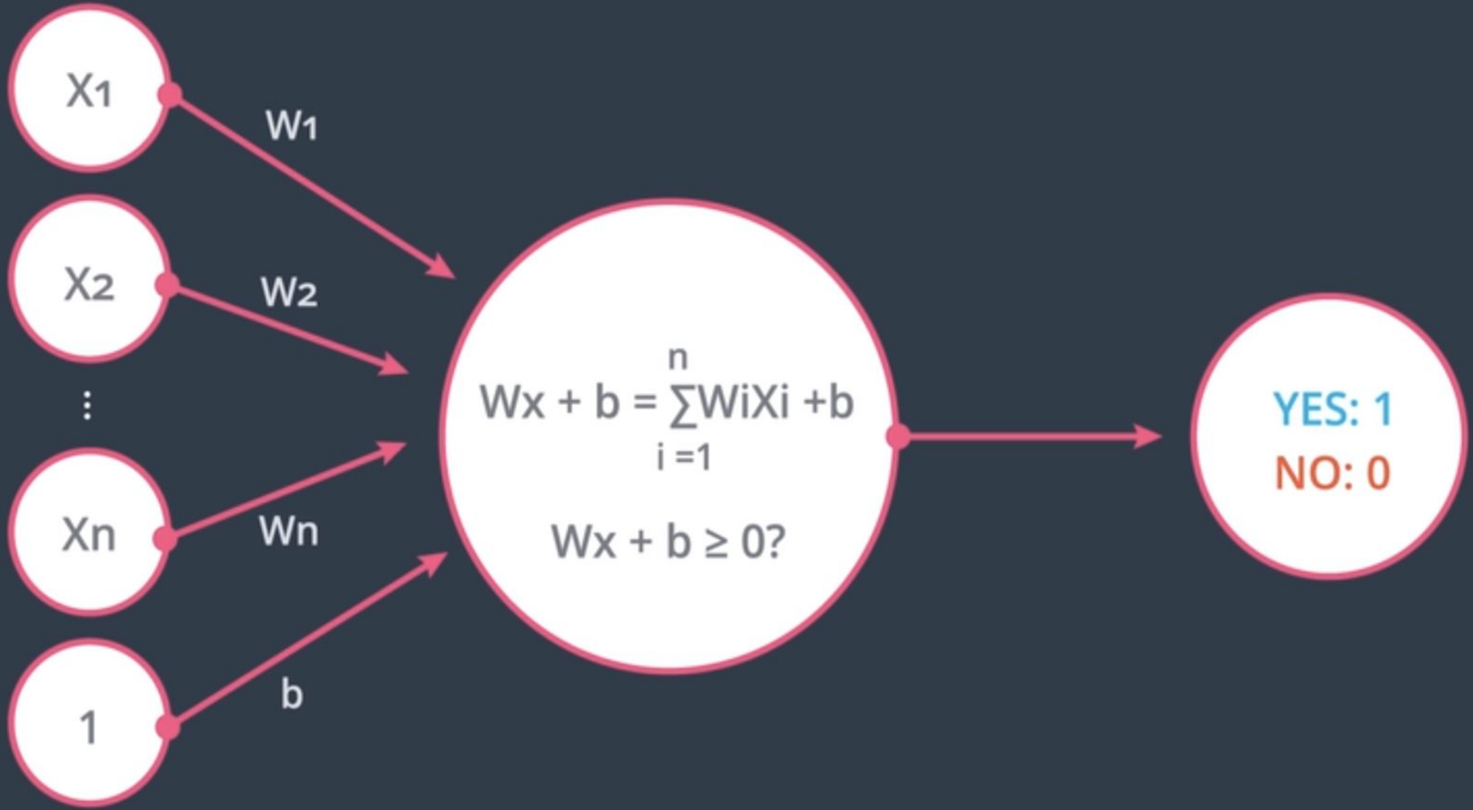
Score=
2*Test + 1*Grades -18

PREDICTION:

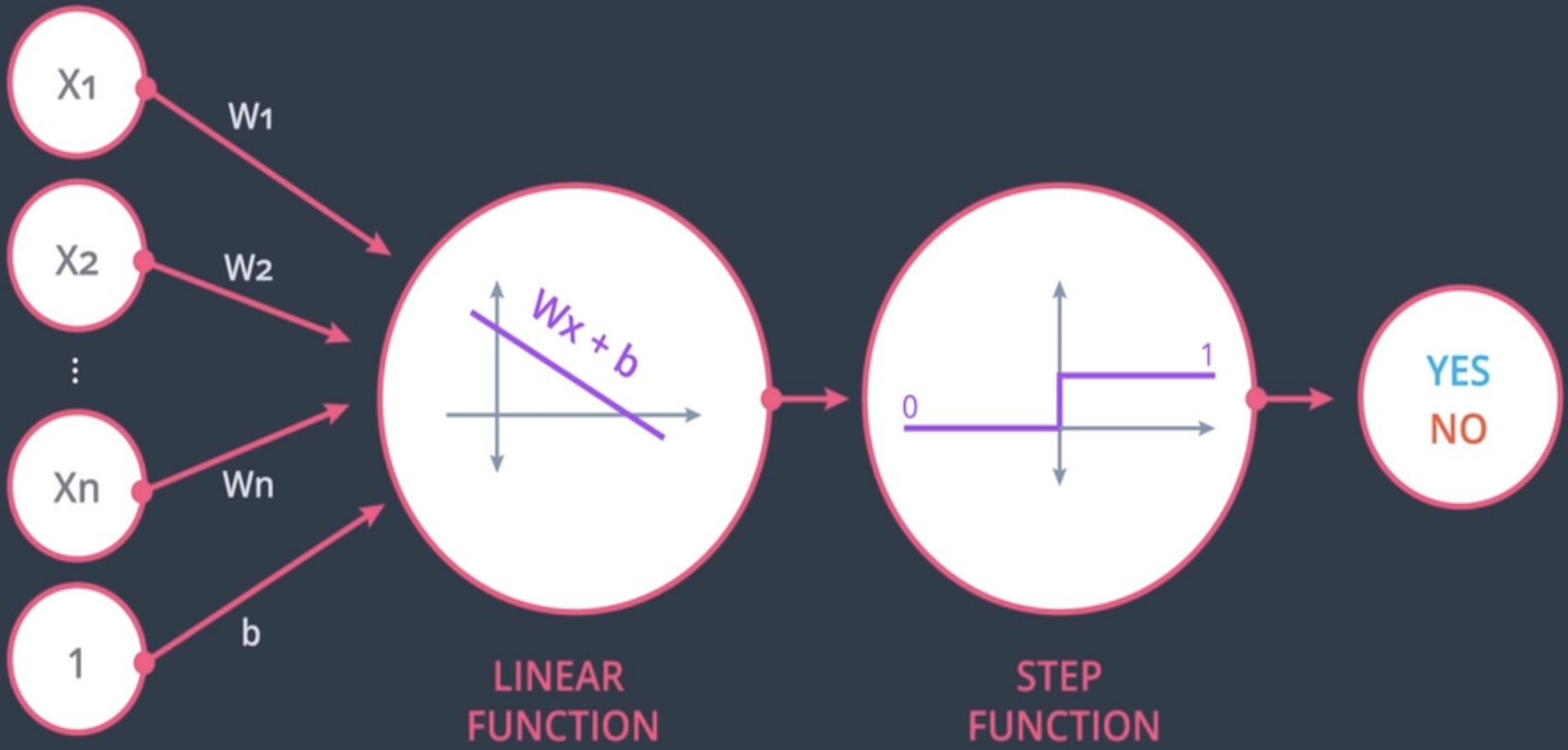
Score ≥ 0 Accept

Score < 0 Reject

Perceptron

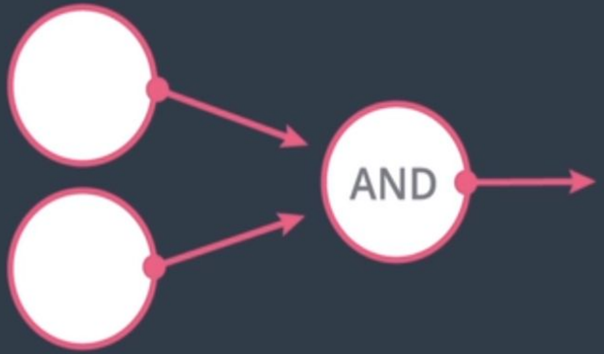


Perceptron

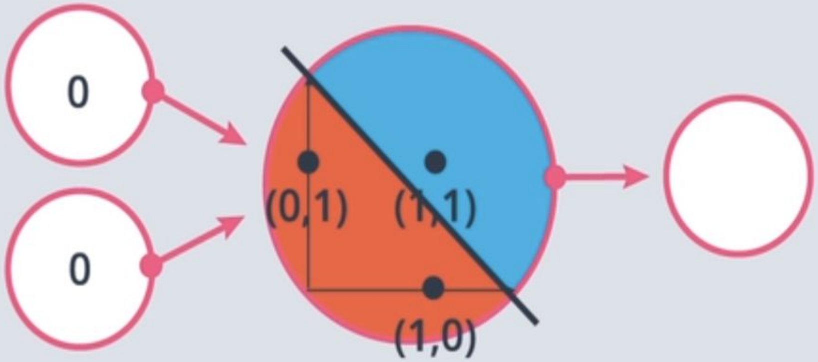


X1
X2
Xn
1
W1
W2
Wn
b
Wx + b
LINEAR FUNCTION
0
1
STEP FUNCTION
YES
NO

AND Perceptron

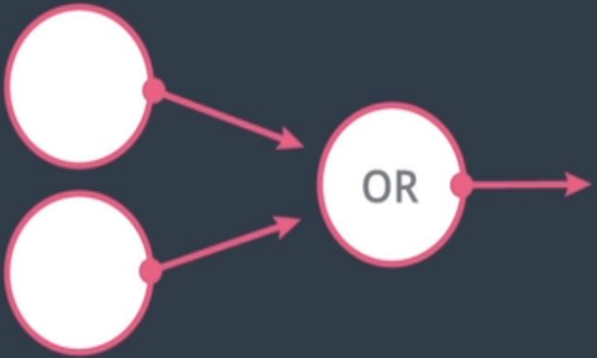


IN	OUT	IN
✓	✓	✓
✓	✗	✗
✗	✓	✗
✗	✗	✗

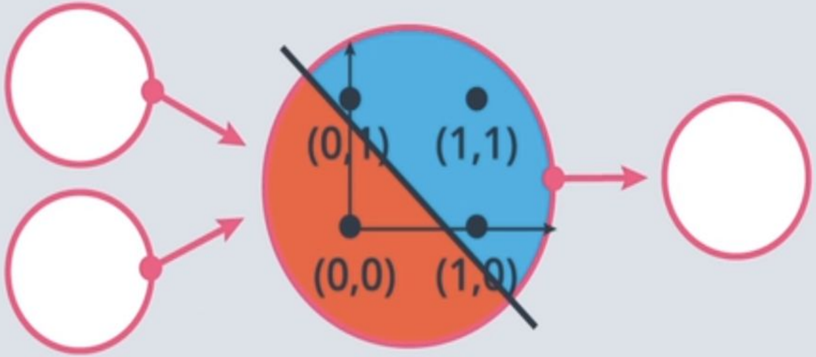


IN	IN	OUT
1	1	1
1	0	0
0	1	0
		0

OR Perceptron

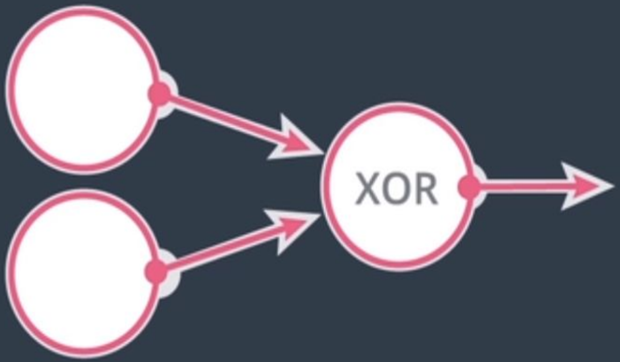


IN	OUT	IN
✓	✓	✓
✓	✗	✓
✗	✓	✓
✗	✗	✗

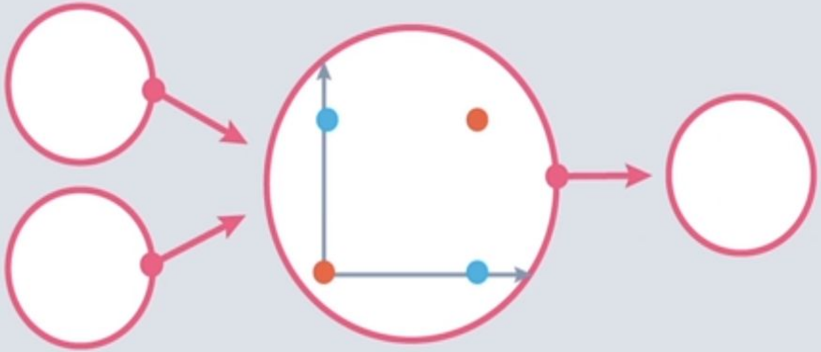


IN	IN	OUT
1	1	1
1	0	1
0	1	1
0	0	0

XOR Perceptron



IN	OUT	IN
✓	✓	✗
✓	✗	✓
✗	✓	✓
✗	✗	✗



IN	IN	OUT
1	1	0
1	0	1
0	1	1
0	0	0

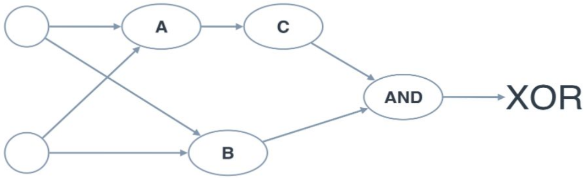
IN	IN	OUT
1	1	0
1	0	1
0	1	1
0	0	0

Quiz: Build an XOR Multi-Layer Perceptron

Now, let's build a multi-layer perceptron from the AND, NOT, and OR perceptrons to create XOR logic!

The neural network below contains 3 perceptrons, A, B, and C. The last one (AND) has been given for you. The input to the neural network is from the first node. The output comes out of the last node.

The multi-layer perceptron below calculates XOR. Each perceptron is a logic operation of AND, OR, and NOT. However, the perceptrons A, B, and C don't indicate their operation. In the following quiz, set the correct operations for the perceptrons to calculate XOR.



QUIZ QUESTION

Set the operations for the perceptrons in the XOR neural network.

These are the correct matches.

Perceptron	Operators
A	AND
B	OR
C	NOT

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Man, works at a factory. What app do we recommend?

- ○ Pokémon Go
- ○ WhatsApp
- ○ Snapchat

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Girl, goes to high school.
What app do we recommend?

- ○ Pokémon Go
- ○ WhatsApp
- ○ Snapchat

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Woman, works at an office.
What app do we recommend?

- Pokémon Go
- WhatsApp
- Snapchat

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Man, works at a factory.
What app do we recommend?

- ○ Pokémon Go
- ○ WhatsApp
- ● Snapchat

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Between **Gender** and **Occupation**, which one seems more decisive for predicting what app will the users download?

- ○ **Gender**
- ○ **Occupation**

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Between **Gender** and **Occupation**, which one seems more decisive for predicting what app will the users download?

- ○ **Gender**
- ○ **Occupation**

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Between **Gender** and **Occupation**, which one seems more decisive for predicting what app will the users download?

- ○ Gender
- ○ Occupation

Recommending Apps

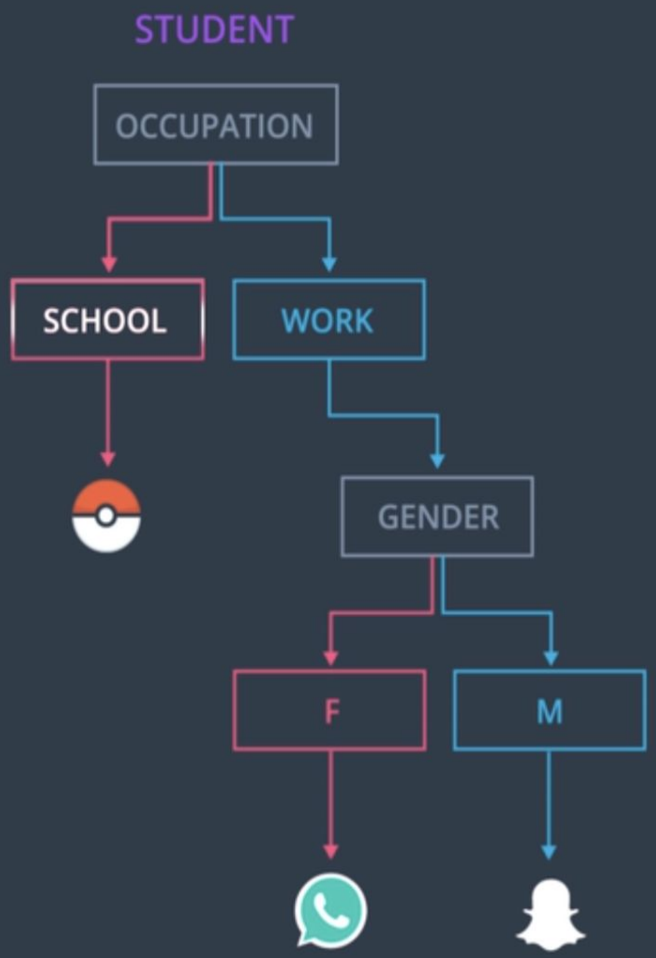
Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Quiz: Between **Gender** and **Occupation**, which one seems more decisive for predicting what app will the users download?

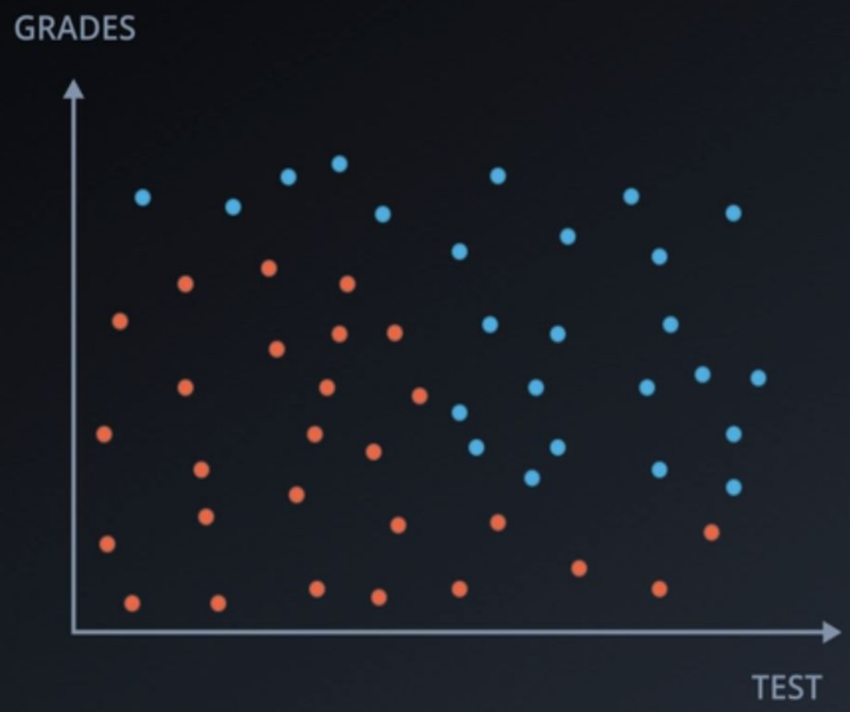
- Gender
- Occupation

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	



Student Admissions



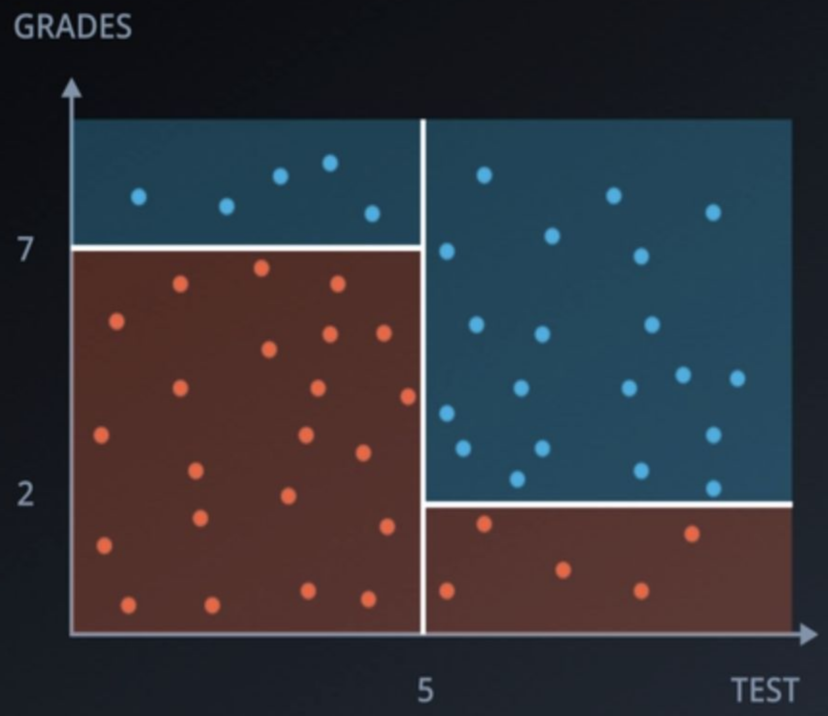
Quiz: Between grades and test, which one determines student acceptance better?

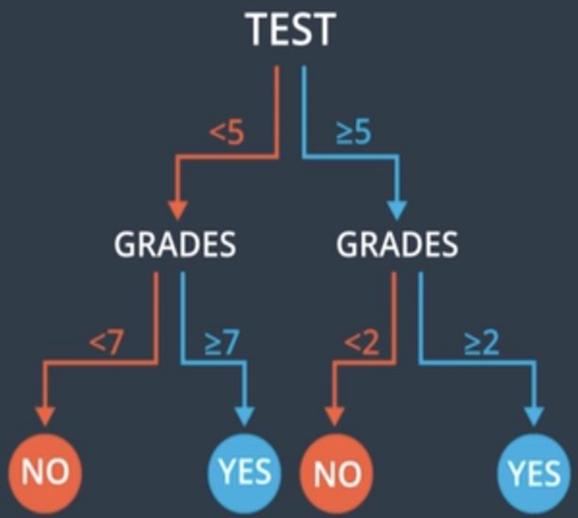
Student Admissions

Quiz: Between a horizontal and a vertical line, which one would cut the data better?

- ○ Horizontal
- ● Vertical

Student Admissions





Entropy

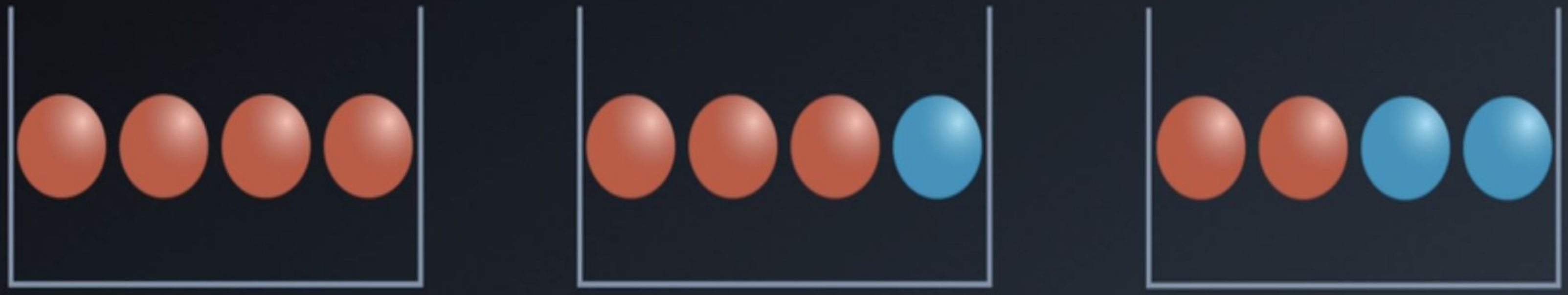
Entropy

Low

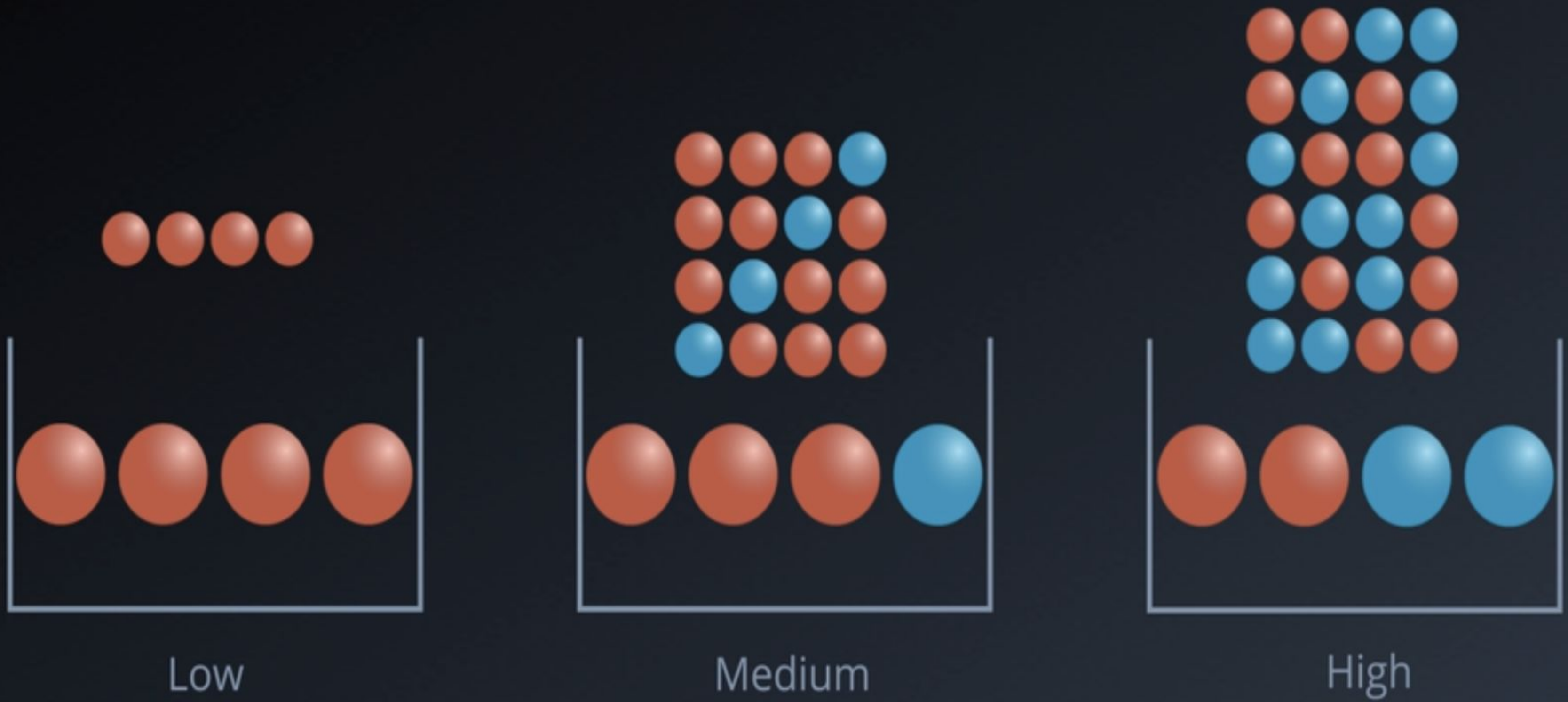
Medium

High

Entropy



Entropy



Entropy

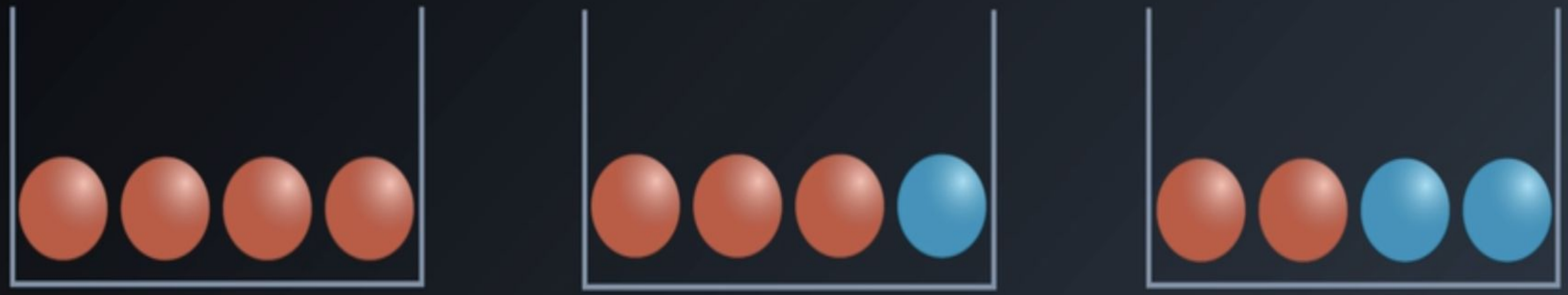
Entropy


HIGH KNOWLEDGE
Low Entropy
MEDIUM KNOWLEDGE
Medium Entropy
LOW KNOWLEDGE
High Entropy

Game

Quiz

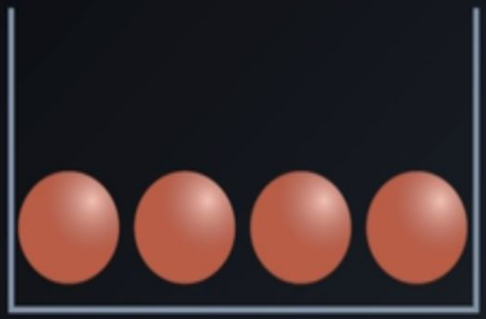
Which one is the best bucket to play the game with?
Which one is the worst?



Quiz

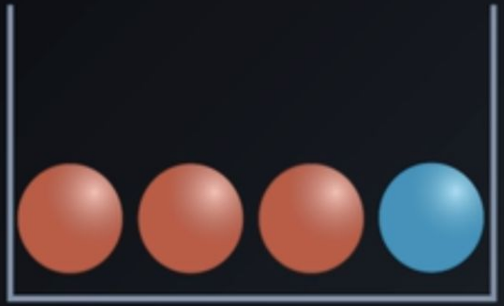
Which one is the best bucket to play the game with?
Which one is the worst?

Game



$$1 \times 1 \times 1 \times 1 = 1$$

Game



$$0.75 \times 0.75 \times 0.75 \times 0.25 = 0.105$$

Probability of Winning

	P (red)	P $(blue)$	P $(winning)$
	1	1	$1 \times 1 \times 1 \times 1 \boxed{= 1}$
	0.75	0.25	$0.75 \times 0.75 \times 0.75 \times 0.25 \boxed{= 0.105}$
	0.5	0.5	$0.5 \times 0.5 \times 0.5 \times 0.5 \boxed{= 0.0625}$

Products

$0.75 * 0.75 * 0.75 * 0.25 = 0.105$

Quiz:

What function to use?

- ○ sin
- ○ cos
- ○ log
- ○ exp

Products

$$0.75 * 0.75 * 0.75 * 0.25 = 0.105$$

Quiz:

What function to use?

- ○ sin
- ○ cos
- ● log
- ○ exp

$$log(ab) = log(a) + log(b)$$

Entropy

$$Entropy = -\frac{5}{8}log_2\left(\frac{5}{8}\right) - \frac{3}{8}log_2\left(\frac{3}{8}\right) = 0.9544$$

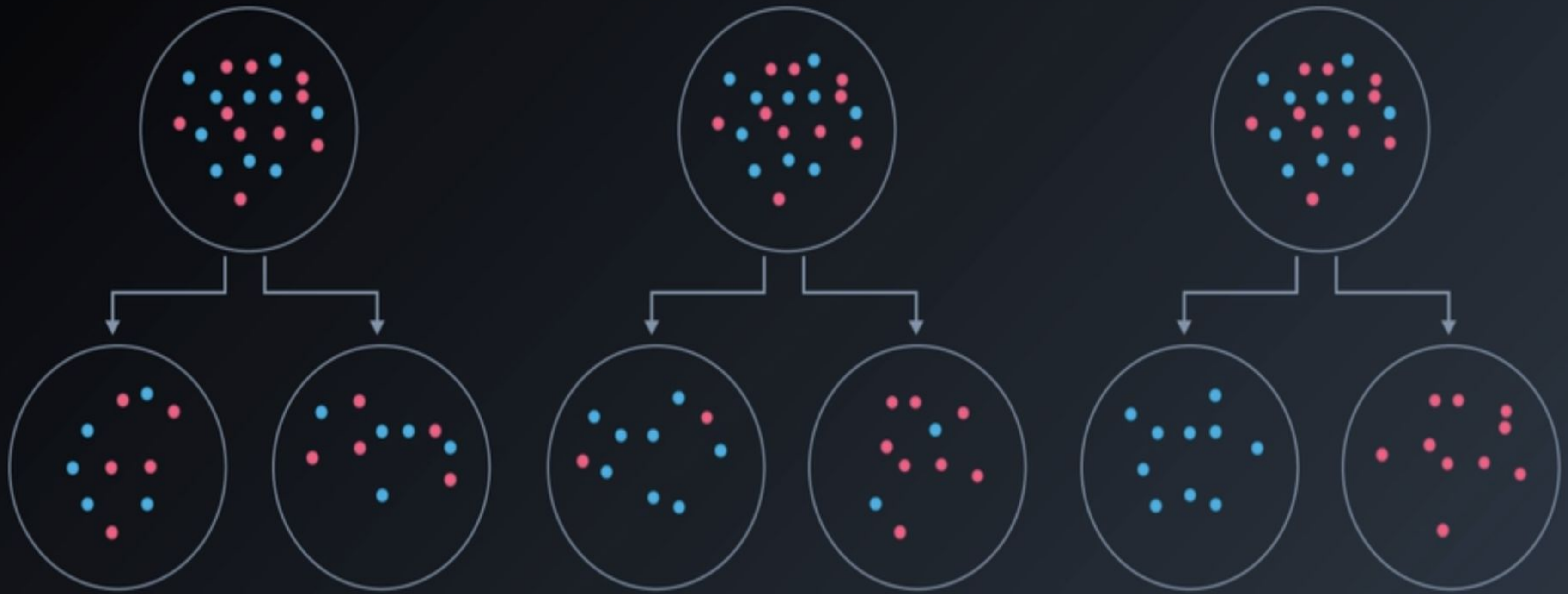
Entropy

$$Entropy = -\frac{m}{m-n} log_2\left(\frac{m}{m+n}\right) - \frac{n}{m+n} log_2\left(\frac{n}{m+n}\right)$$

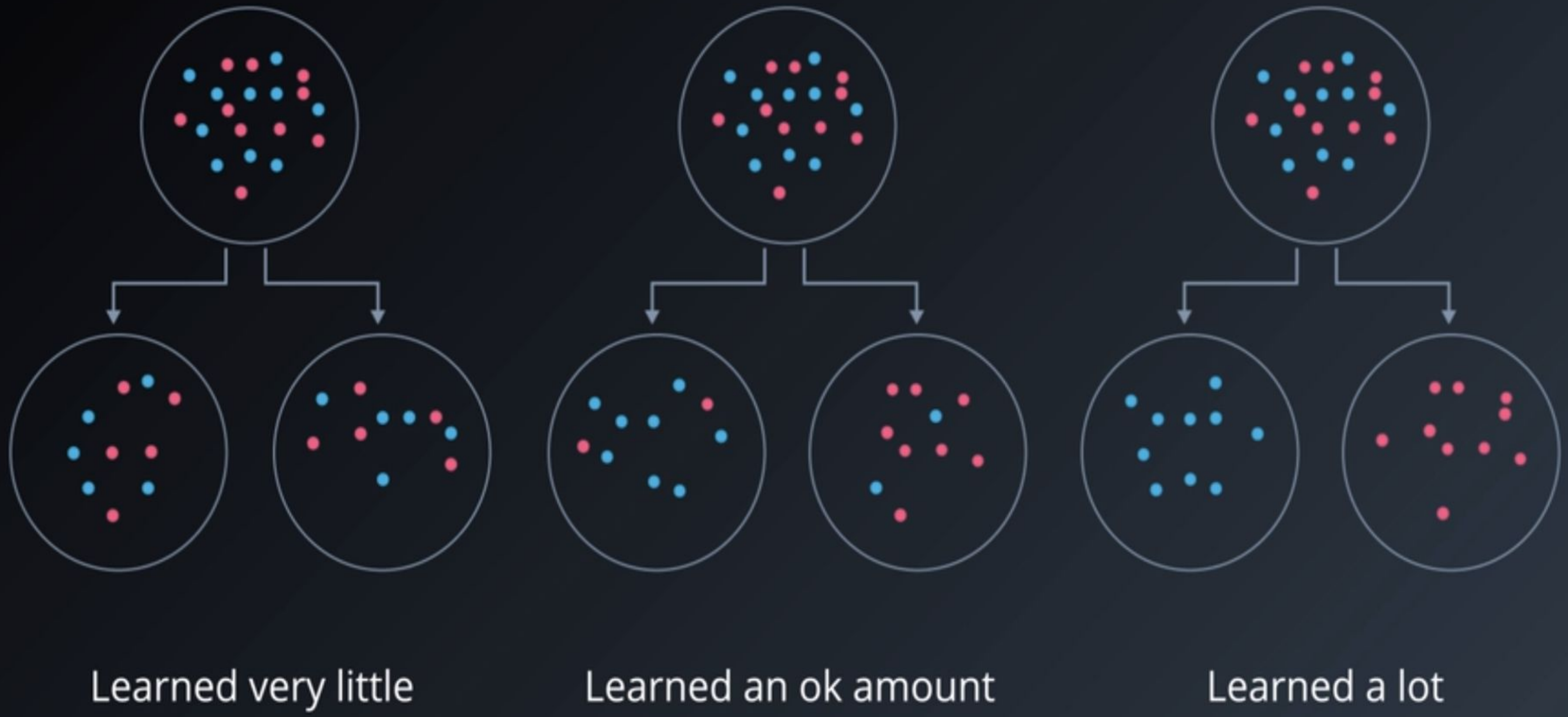
Information Gain

Quiz: Where did we gain more information? Where did we gain less?

Information Gain



Information Gain

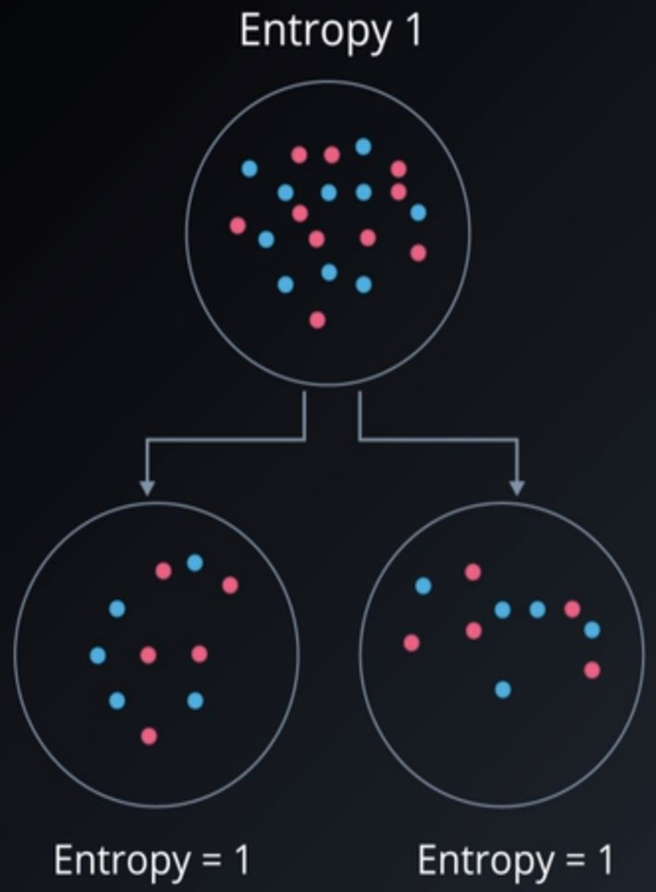


Information Gain =

Entropy(parent)

- 0.5 [Entropy(child 1) + Entropy(child 2)]

Information gain
1-0.72 = 0.28



Information gain
1 - 1 = 0

Information gain
1 - 0 = 1

Information gain = 0 | Information gain = 0.28 | Information gain = 1

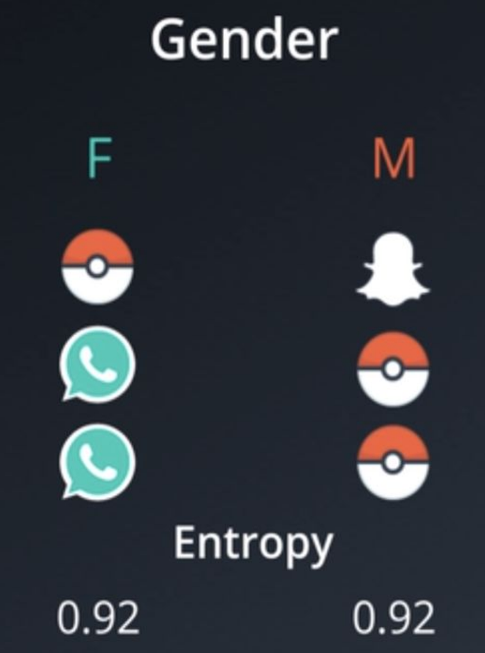
Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

$$\text{Entropy} = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{1}{6}\log_2\left(\frac{1}{6}\right)$$

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	



Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Gender

F	M

Entropy

0.92	0.92

Information gain = 1.46 - 0.92 = 0.54

Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

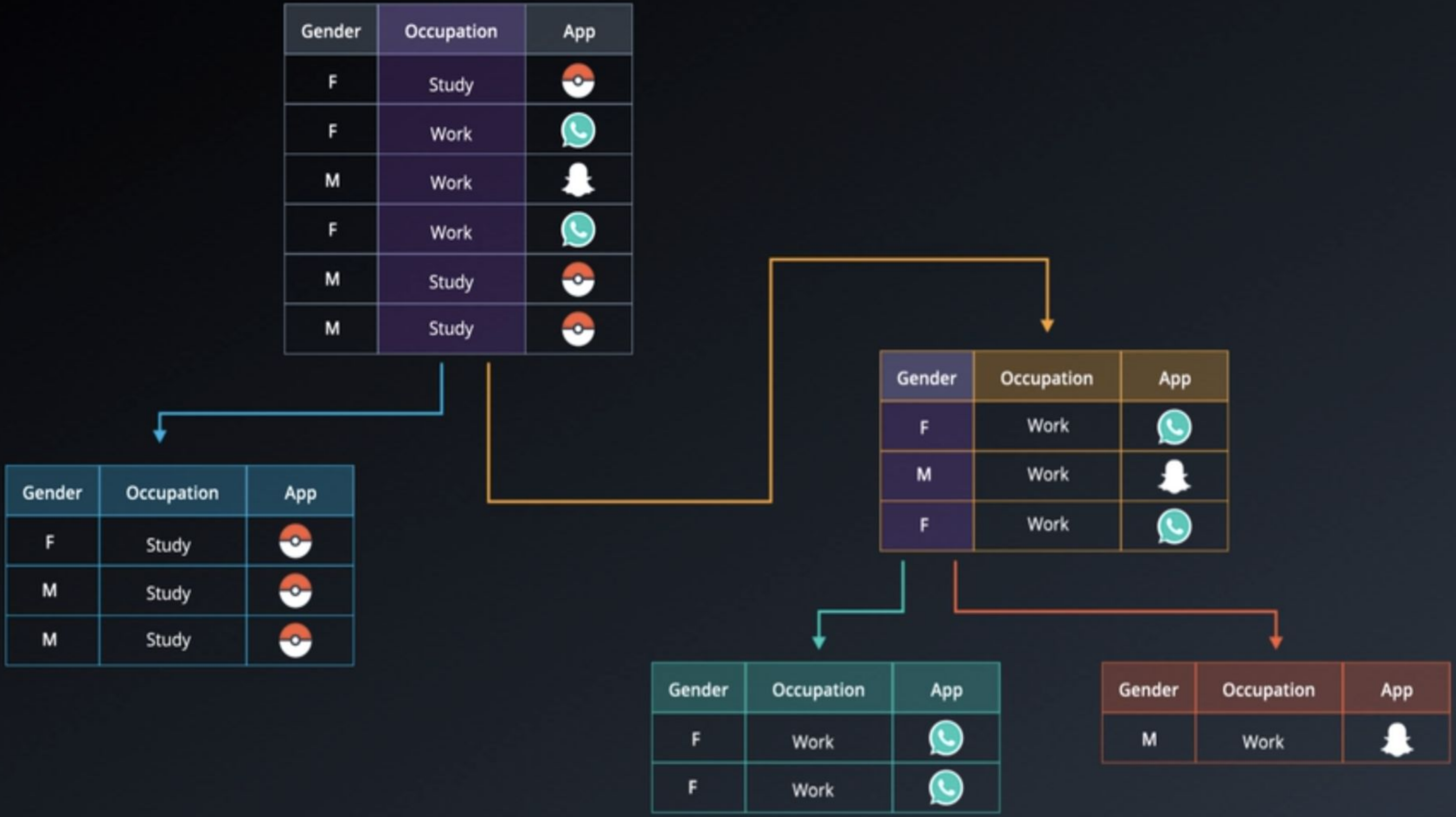
Occupation

S	M

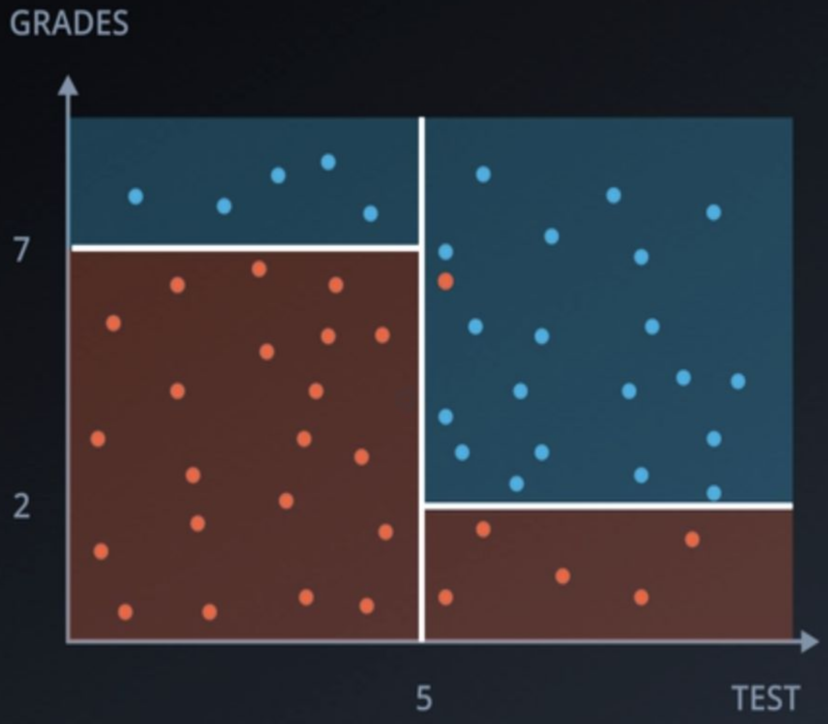
Entropy

S	M
0	0.92

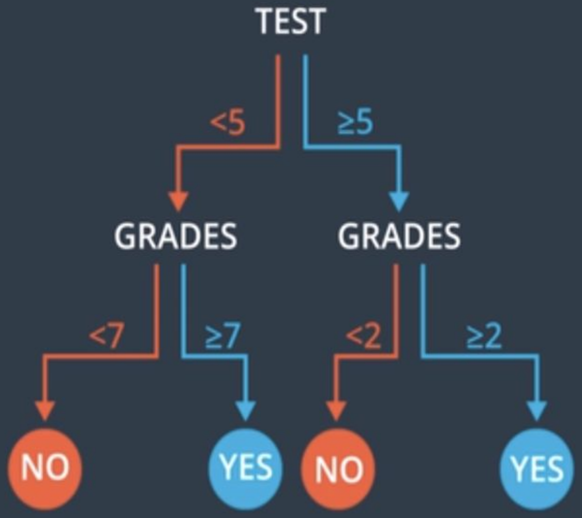
Information gain = 1.46 - 0.46 = 1


Gender
Occupation
App
F
Study
F
Work
M
Work
F
Work
M
Study
M
Study
Gender
Occupation
App
F
Study
M
Study
M
Study
Gender
Occupation
App
F
Work
M
Work
F
Work
Gender
Occupation
App
F
Work
F
Work
Gender
Occupation
App
M
Work

Student Admissions



Question: Out of all of the vertical and horizaontal lines that cut this data, which one maximizes the information gain?



BAYES THEOREM

Alex $P(\text{Alex}) = 0.5$

Brenda $P(\text{Brenda}) = 0.5$

Person had a red sweater

Alex wears red 2 times a week

Brenda wears red 3 times a week

BAYES THEOREM

Alex

Brenda

$P(\text{Alex}) = 0.5$ $P(\text{Brenda}) = 0.5$

$P(\text{Alex}) = 0.4$ $P(\text{Brenda}) = 0.6$

Person had a red sweater

Alex wears red 2 times a week

Brenda wears red 3 times a week

40%

60%

BAYES THEOREM

	Alex	Brenda
Prior	$P(\text{Alex}) = 0.5$	$P(\text{Brenda}) = 0.5$
Posterior	$P(\text{Alex}) = 0.4$	$P(\text{Brenda}) = 0.6$

Person had a red sweater

Alex wears red 2 times a week

Brenda wears red 3 times a week

40% 60%

BAYES THEOREM

Known	Inferred
Probability that Alex wears red	Probability that the person wearing red is Alex
Probability that Brenda wears red	Probability that the person wearing red is Brenda

BAYES THEOREM

Known	Inferred
$P(A)$ $P(R \mid A)$	$P(A \mid R)$

BAYES THEOREM

Alex comes to the office 3 days a week

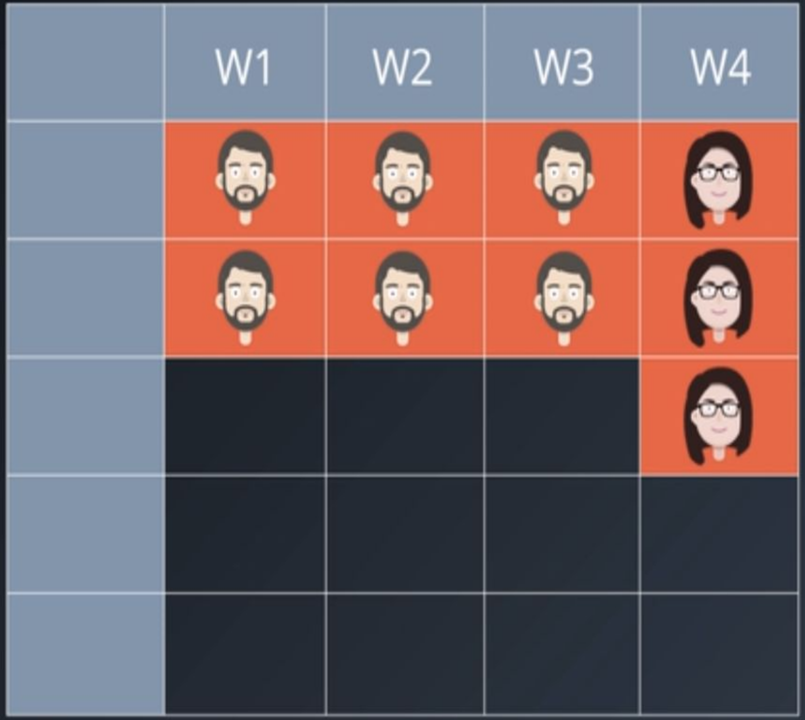
Brenda comes to the office 1 day a week

Prior $P(\text{Alex}) = 0.75$ $P(\text{Brenda}) = 0.25$

Person had a red sweater

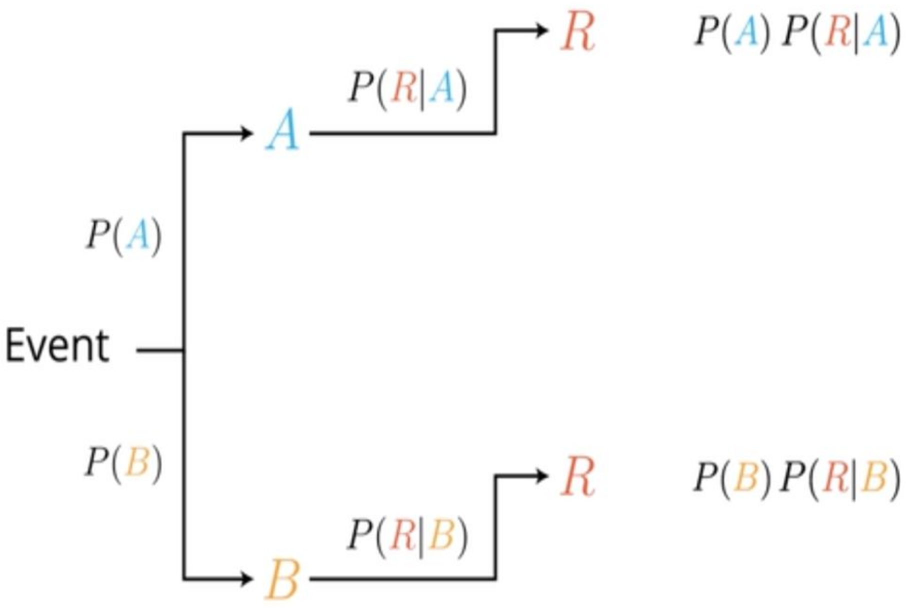
Alex wears red 2 times a week

Brenda wears red 3 times a week



BAYES THEOREM

Prior



Posterior

$$P(A|R) = \frac{P(A)P(R|A)}{P(A)P(R|A) + P(B)P(R|B)}$$

$$P(B|R) = \frac{P(B)P(R|B)}{P(A)P(R|A) + P(B)P(R|B)}$$

BAYES THEOREM

$$P(S|+) = \frac{P(S)P(+|S)}{P(S)P(+|S) + P(H)P(+|H)}$$

$$= \frac{0.0001 * 0.99}{0.0001 * 0.99 + 0.9999 * 0.01}$$

$$= 0.0098$$

$$< 1\%$$

S: sick

H: healthy

+ : positive

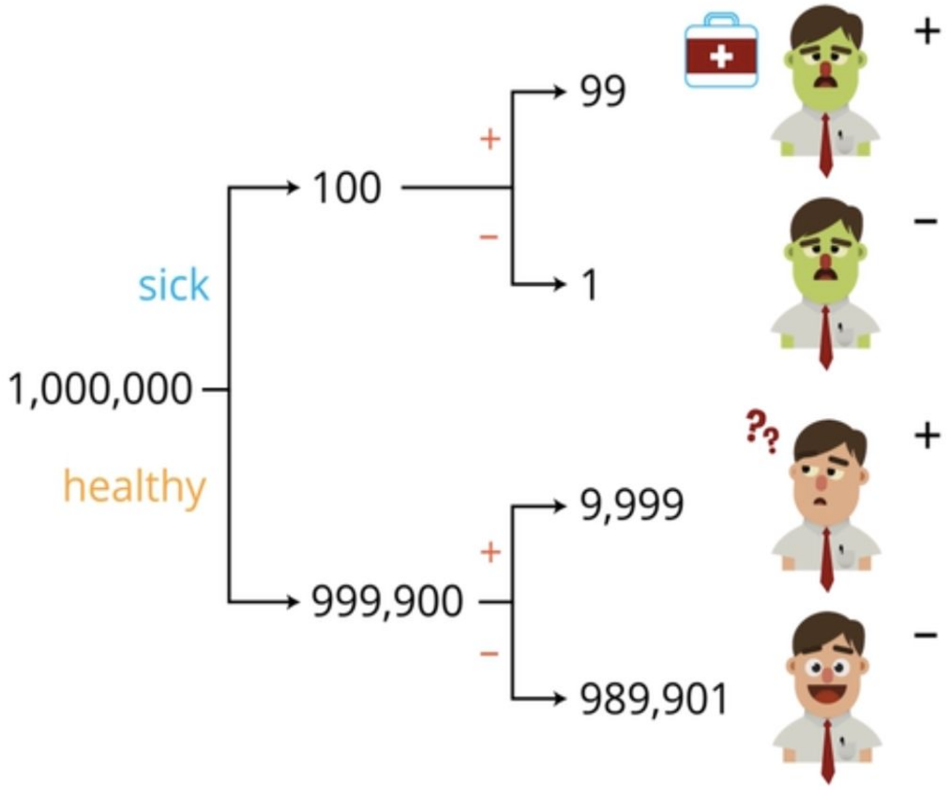
$$P(S) = 0.0001$$

$$P(H) = 0.9999$$

$$P(+|S) = 0.99$$

$$P(+|H) = 0.01$$

BAYES THEOREM

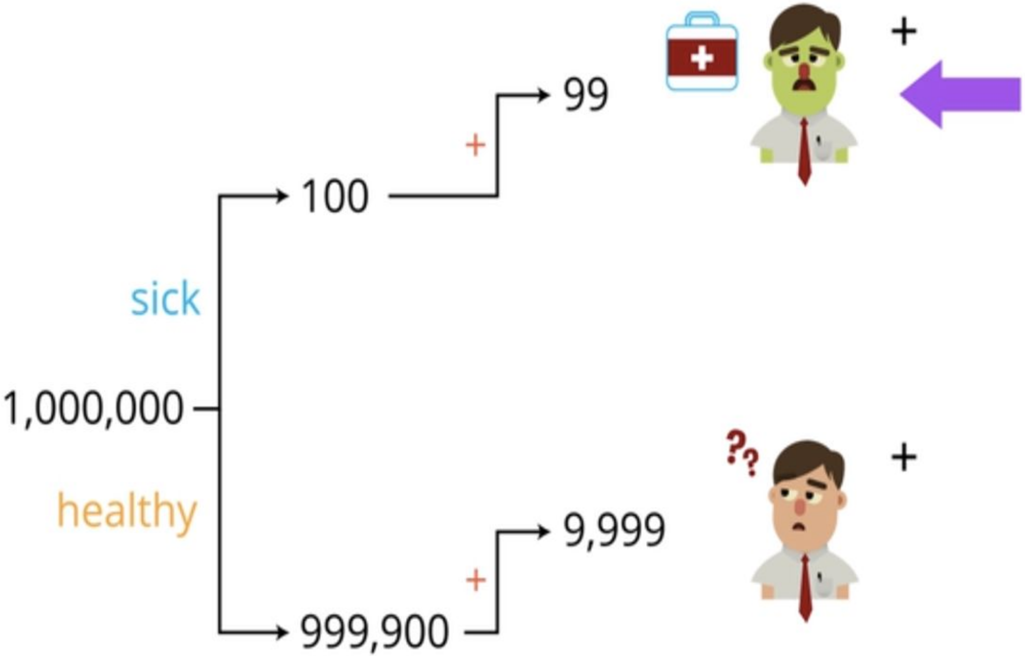


1 out of every 10,000 patients is sick

Test has 99% accuracy

Patient tested positive

BAYES THEOREM



1 out of every 10,000 patients is sick

Test has 99% accuracy

Patient tested positive

$$P(\text{sick}|+) = \frac{99}{9,999 + 99}$$

$$= 0.0098$$

$$< 1$$

FALSE POSITIVES

1% errors

0.01% sick

FALSE POSITIVES

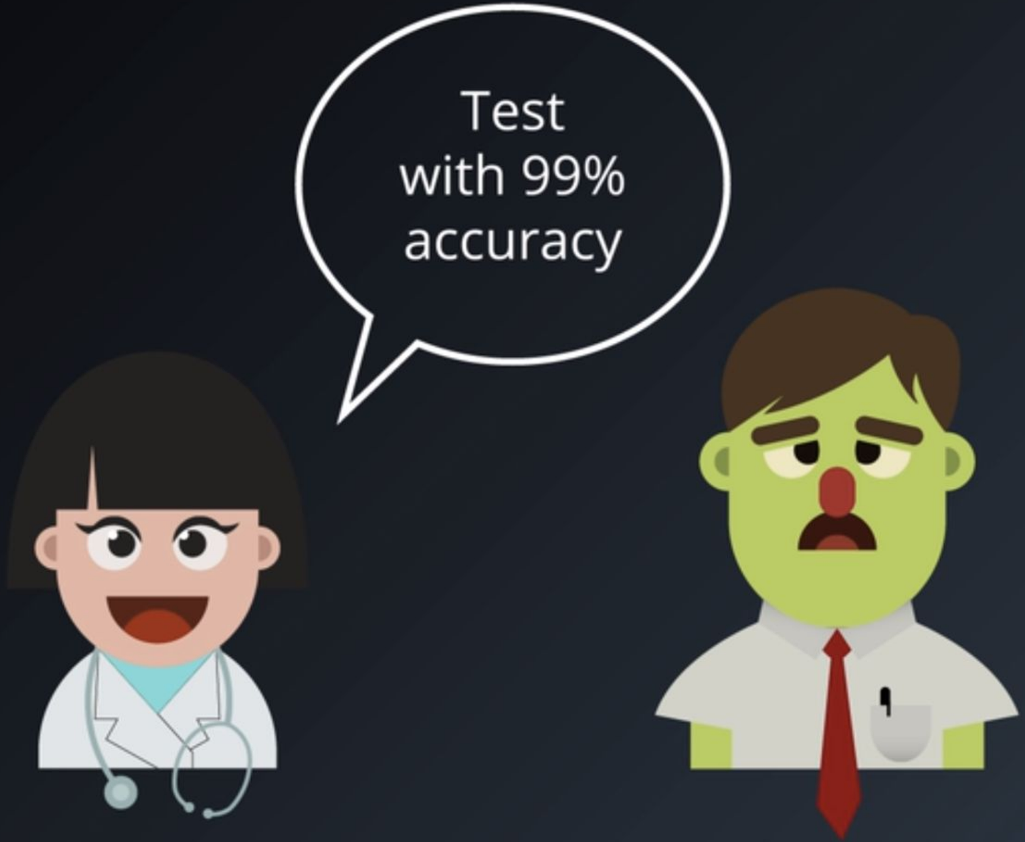
1% errors

100 out of every 10,000

0.01% sick

1 out of every 10,000

BAYES THEOREM



BAYES THEOREM

BAYES THEOREM

99% Accuracy
1 out of 10,000 people are sick

Quiz:
What is the probability of being sick?

BAYES THEOREM

99% Accuracy
1 out of 10,000
people are sick

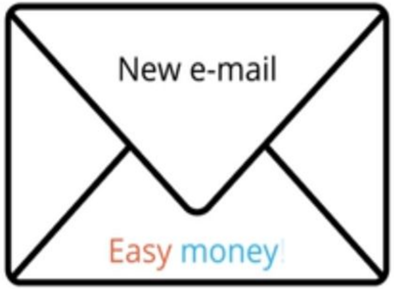
Quiz:
What is the probability
of being sick?

- ◯ 0%-20%
- ◯ 20%-40%
- ◯ 40%-60%
- ◯ 60%-80%
- ◯ 80%-100%

NAIVE BAYES

Spam
Win money now! Make cash easy! Cheap money, reply.
Ham
How are you? There you are! Can I borrow money? Say hi to grandma. Was the exam easy?

NAIVE BAYES



Spam
Win money now!
Make cash easy!
Cheap money, reply.

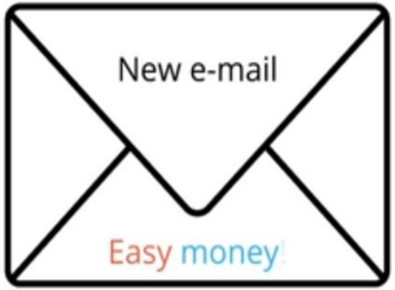
Ham
How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

Quiz

What's the probability that an e-mail contains the word 'easy', given that it is spam?

- ○ 1/5
- ○ 1/4
- ○ 1/3
- ○ 2/3
- ○ 4/5

NAIVE BAYES



Spam
Win money now!
Make cash easy!
Cheap money, reply.

Ham
How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

Quiz

What's the probability that an e-mail contains the word 'money', given that it is spam?

- 1/5
- 1/4
- 1/3
- 2/3
- 4/5

NAIVE BAYES

Spam

Win money now!
Make cash easy!
Cheap money, reply.

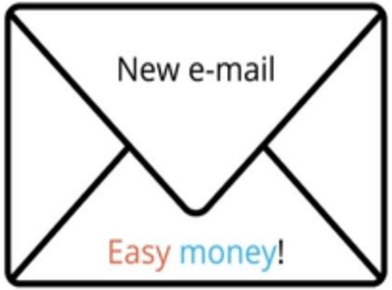
Ham

How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

P ('easy' | spam) = 1/3

P ('money' | spam) = 2/3

NAIVE BAYES



Spam
Win money now!
Make cash easy!
Cheap money, reply.

Ham
How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

$P(\text{'easy'} \mid \text{spam}) = 1/3$

$P(\text{'money'} \mid \text{spam}) = 2/3$

$P(\text{'easy'} \mid \text{ham}) = 1/5$

$P(\text{'money'} \mid \text{ham}) = 1/5$

NAIVE BAYES

Known

Spam

Win money now!

Make cash easy!

Cheap money, reply.

Ham

How are you?

There you are!

Can I borrow money?

Say hi to grandma.

Was the exam easy?

$P(\text{'easy'} \mid \text{spam}) = 1/3$

$P(\text{'money'} \mid \text{spam}) = 2/3$

Inferred

Contains "easy"

Make cash easy!

Was the exam easy?

Contains "money"

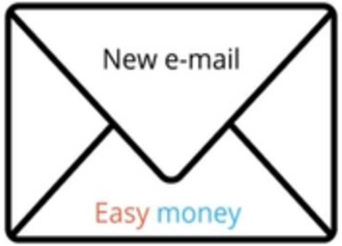
Win money now!

Cheap money, reply.

Can I borrow money?

$P(\text{spam} \mid \text{'easy'}) = 1/2$

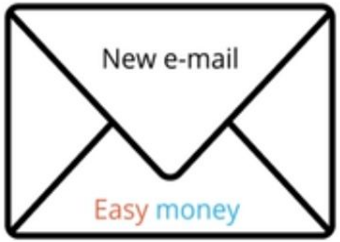
NAIVE BAYES



Spam Win money now! Make cash easy! Cheap money, reply.
Ham How are you? There you are! Can I borrow money? Say hi to grandma. Was the exam easy?

P (spam | 'easy')

NAIVE BAYES



Spam

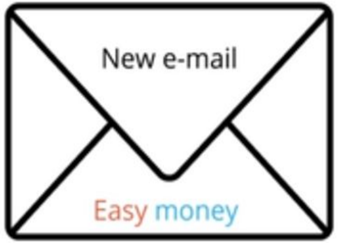
Win money now!
Make cash easy!
Cheap money, reply.

Ham

How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

P (spam | 'easy')

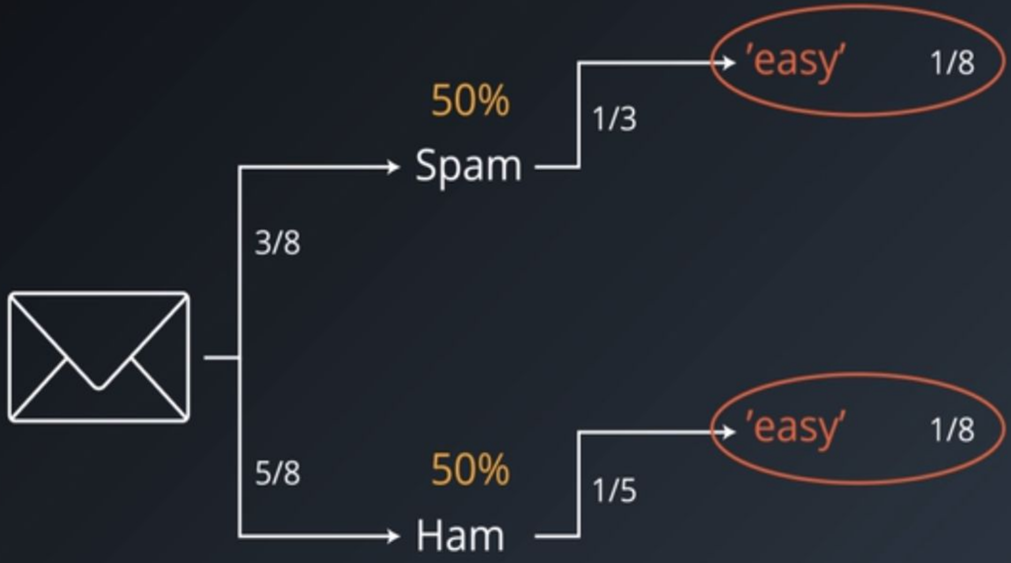
NAIVE BAYES



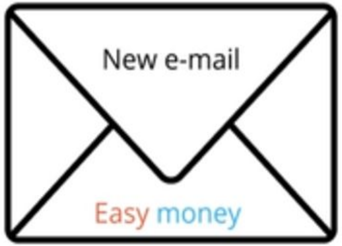
Spam
Win money now!
Make cash easy!
Cheap money, reply.

Ham
How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

P (spam | 'easy')



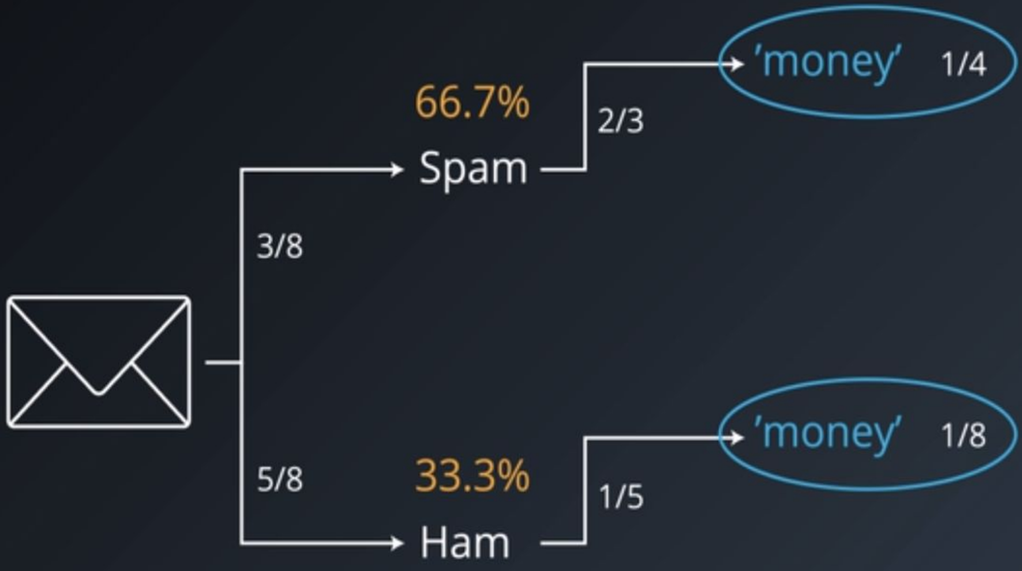
NAIVE BAYES



Spam
Win money now!
Make cash easy!
Cheap money, reply.

Ham
How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

P (spam | 'money')



NAIVE ASSUMPTION

$$P(A \cap B) = P(A)P(B)$$

0   >0   >0

CONDITIONAL PROBABILITY

$$P(A \mid B)\,P(B) = P(B \mid A)\,P(A)$$

CONDITIONAL PROBABILITY

$$P(A \mid B) \propto P(B \mid A)P(A)$$

NAIVE BAYES

$$P(\text{spam} \mid \text{'easy'}, \text{'money'}) \propto P(\text{'easy'}, \text{'money'} \mid \text{spam})\, P(\text{spam})$$

$$P(A \mid B) \propto P(B \mid A) P(A)$$

CONDITIONAL PROBABILITY

NAIVE BAYES

$$P(\text{spam} \mid \text{'easy'}, \text{'money'}) \propto P(\text{'easy'} \mid \text{spam})\, P(\text{'money'} \mid \text{spam})\, P(\text{spam})$$

$$P(A \,\&\, B) = P(A)P(B)$$

NAIVE ASSUMPTION

Spam Win money now! Make cash easy! Cheap money, reply.
Ham How are you? There you are! Can I borrow money? Say hi to grandma. Was the exam easy?

NAIVE BAYES

$$\underset{1/12}{P(\text{spam} \mid \text{'easy'}, \text{'money'})} \propto \underset{1/3}{P(\text{'easy'} \mid \text{spam})}\ \underset{2/3}{P(\text{'money'} \mid \text{spam})}\ \underset{3/8}{P(\text{spam})}$$

$$P(\text{ham} \mid \text{'easy'}, \text{'money'}) \propto \underset{1/5}{P(\text{'easy'} \mid \text{ham})}\ \underset{1/5}{P(\text{'money'} \mid \text{ham})}\ \underset{5/8}{P(\text{ham})}$$

NAIVE BAYES

$$P(\text{spam} \mid \text{'easy'}, \text{'money'}) \propto \frac{1}{12} \longrightarrow \boxed{}$$

$$P(\text{ham} \mid \text{'easy'}, \text{'money'}) \propto \frac{1}{40} \longrightarrow \boxed{}$$

NAIVE BAYES

$$P(\text{spam} \mid \text{'easy'}, \text{'money'}) = \frac{1}{12}$$

$$P(\text{ham} \mid \text{'easy'}, \text{'money'}) = \frac{1}{40}$$

NAIVE BAYES

$$P(spam \mid \text{'easy'}, \text{'money'}) \qquad \frac{\frac{1}{12}}{\frac{1}{12} + \frac{1}{40}}$$

$$P(ham \mid \text{'easy'}, \text{'money'})$$

NAIVE BAYES

$$\frac{P(\text{spam} \mid \text{'easy'}, \text{'money'})}{P(\text{ham} \mid \text{'easy'}, \text{'money'})} = \frac{10}{13} \qquad \frac{\frac{1}{40}}{\frac{1}{12} + \frac{1}{40}}$$

NAIVE BAYES

$$P(\text{spam} \mid \text{'easy'}, \text{'money'}) = \frac{10}{13}$$

$$P(\text{ham} \mid \text{'easy'}, \text{'money'}) = \frac{3}{13}$$

NAIVE BAYES

$$P(\text{spam} \mid \text{'easy'}, \text{'money'}, \ldots, \text{'cheap'}) \propto P(\text{'easy'}, \text{'money'}, \ldots, \text{'cheap'} \mid \text{spam})\, P(\text{spam})$$

$$P(\text{'easy'} \mid \text{spam})\, P(\text{'money'} \mid \text{spam}) \cdots P(\text{'cheap'} \mid \text{spam})\, P(\text{spam})$$

NAIVE BAYES

P(spam | 'easy', 'money', ..., 'cheap')

P(ham | 'easy', 'money', ..., 'cheap')

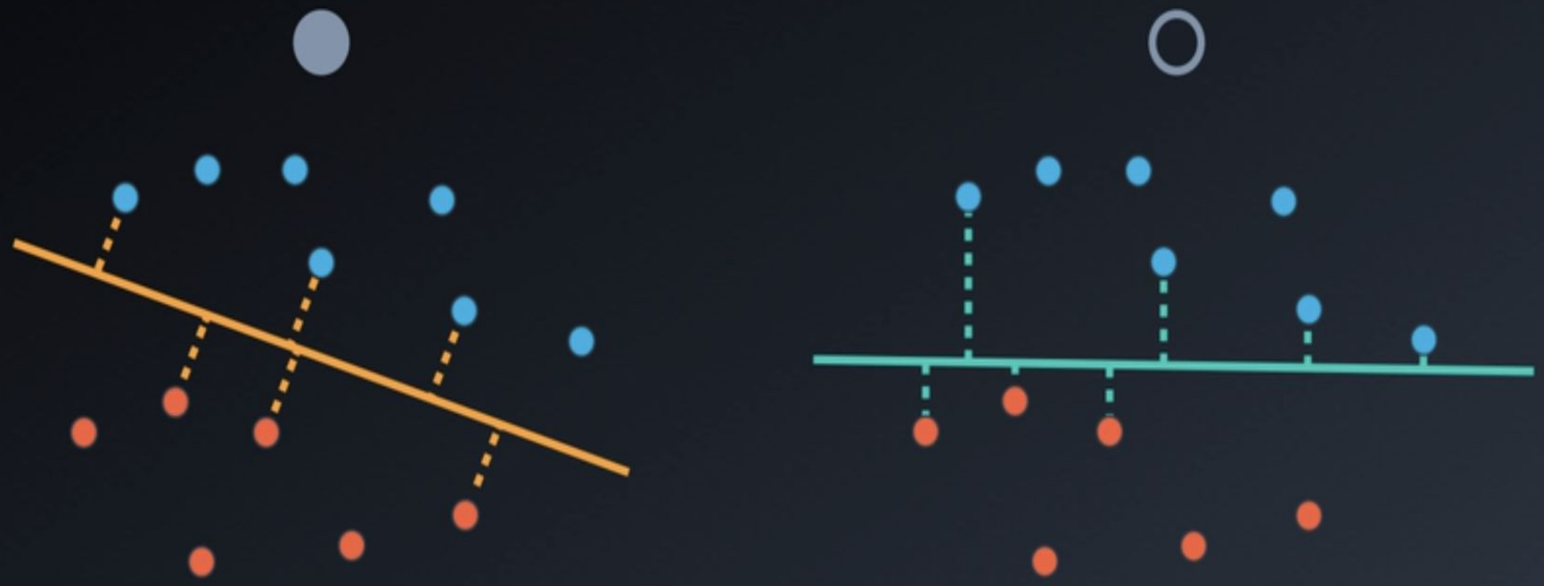
normalize

Probability of spam

Probability of ham

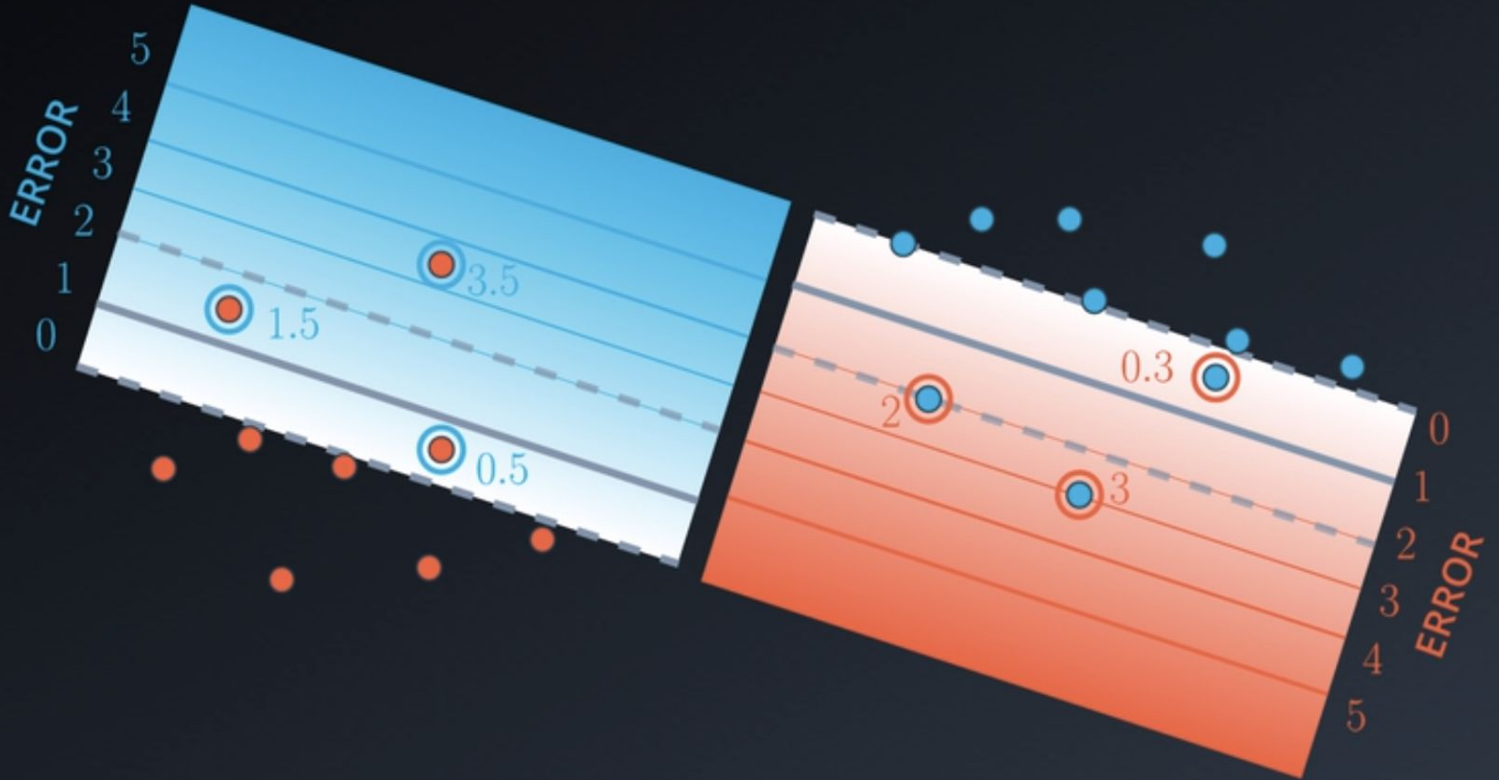
Cutting data with style

Quiz: Which one is a better line?



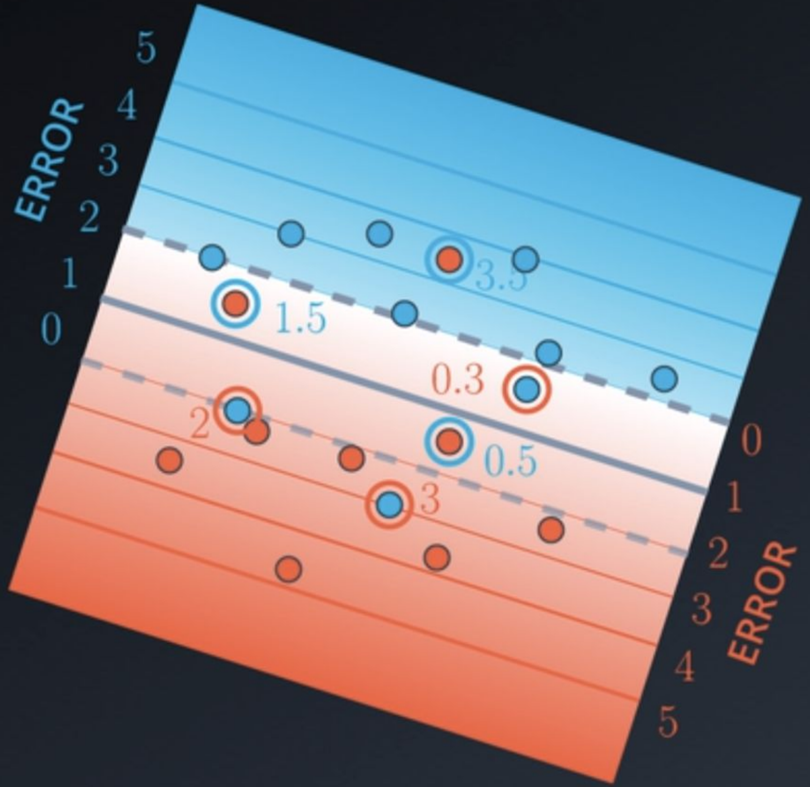
Classification Error

$$Error = 1.5+3.5+0.5+2+3+0.3 = 10.8$$



Classification Error

$$Error = 1.5 + 3.5 + 0.5 + 2 + 3 + 0.3 = 10.8$$

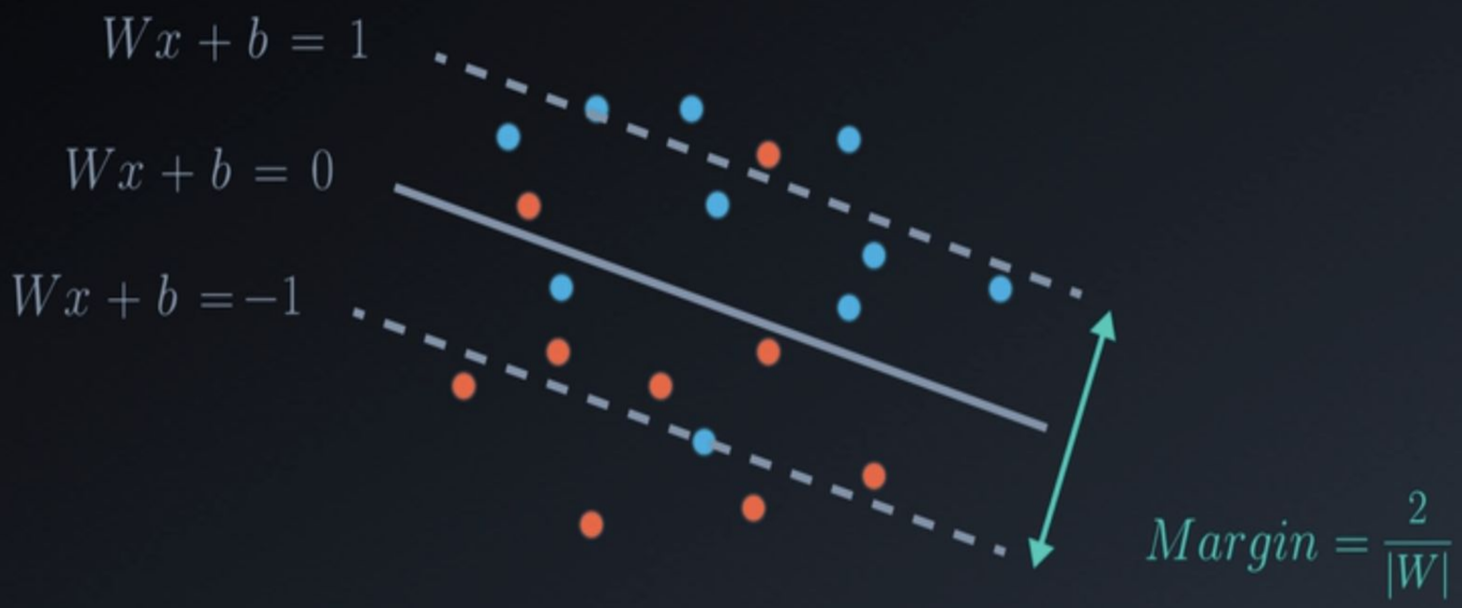


Margin Error

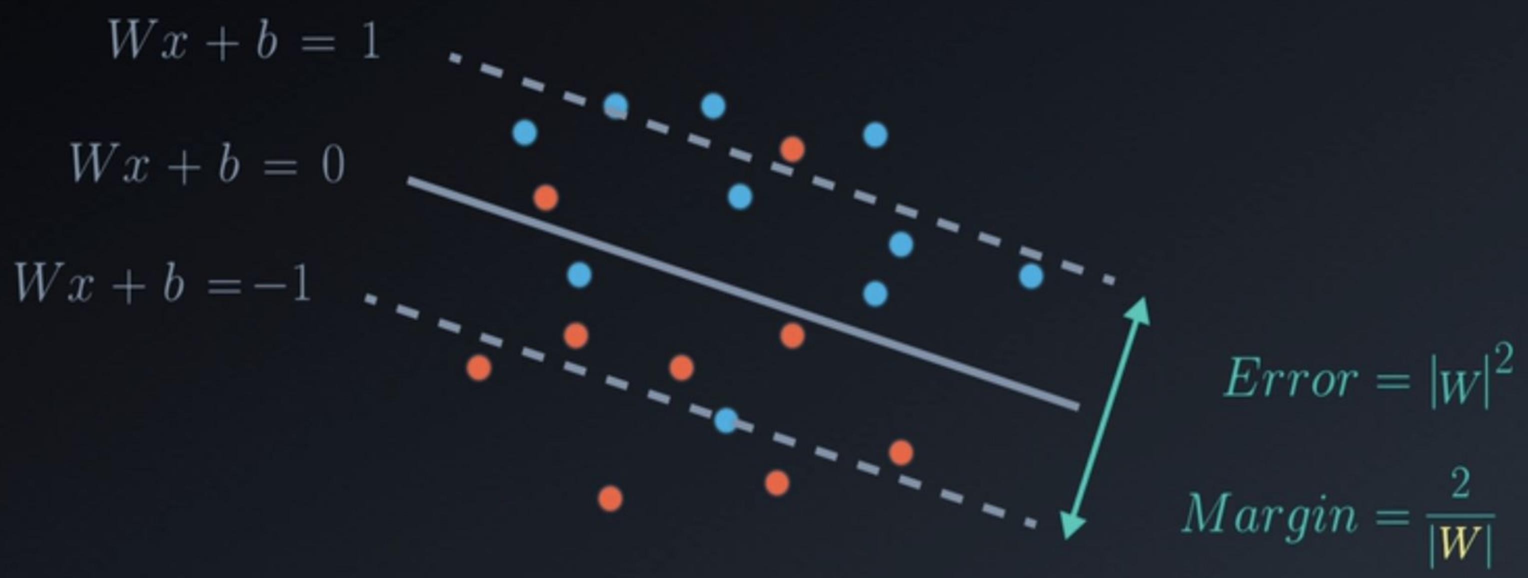
LARGE MARGIN
SMALL ERROR

SMALL MARGIN
LARGE ERROR

Margin Error

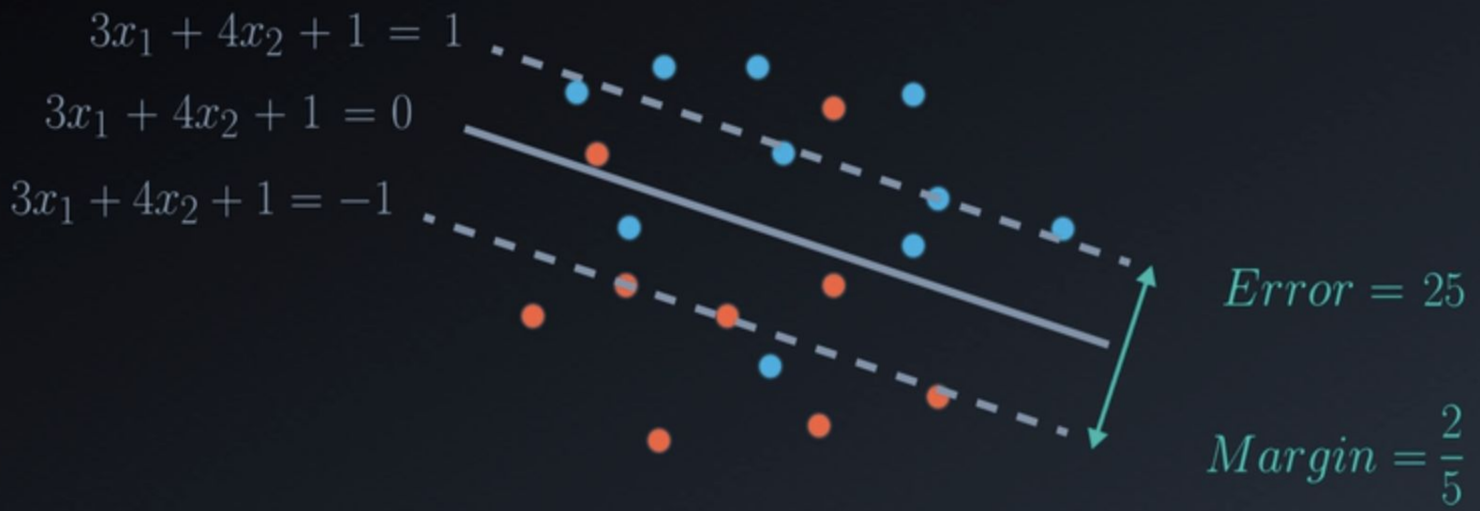


Margin Error



LARGE MARGIN ⟶ SMALL ERROR
SMALL MARGIN ⟶ LARGE ERROR

Example

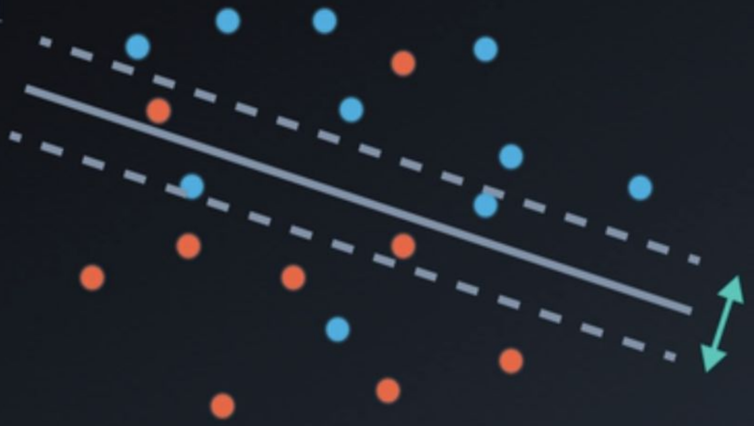


$W = (3, 4)$ $\quad$ $w_1 x_1 + w_2 x_2 + b = 0$ $\quad$ $|W|^2 = 3^2 + 4^2 = 25$

$b = 1$ $\quad$ $3x_1 + 4x_2 + 1 = 0$ $\quad$ $\frac{2}{|W|} = \frac{2}{5}$

Example

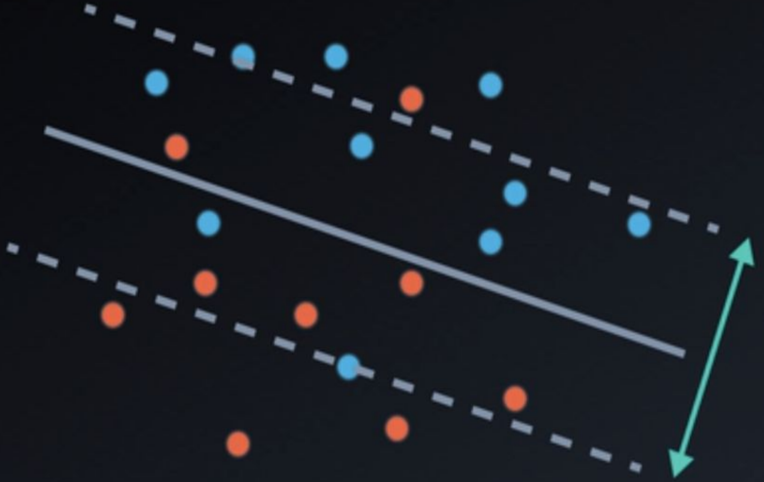
$$6x_1 + 8x_2 + 2 = 1$$
$$6x_1 + 8x_2 + 2 = 0$$
$$6x_1 + 8x_2 + 2 = -1$$



$$Error = 100$$
$$Margin = \frac{2}{10}$$

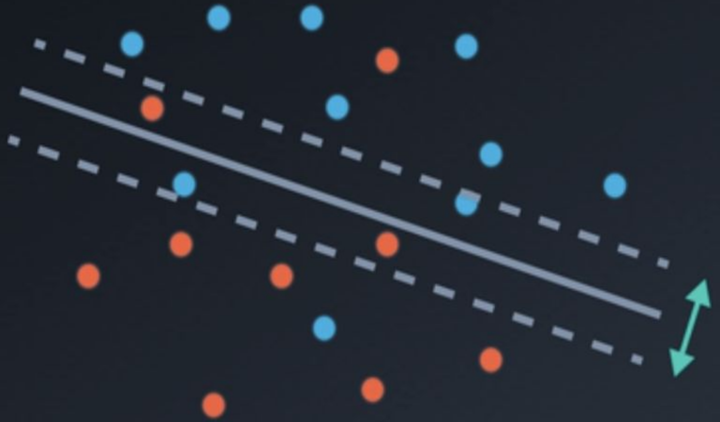
$$W = (6, 8) \qquad w_1x_1 + w_2x_2 + b = 0 \qquad |W|^2 = 6^2 + 8^2 = 100$$

$$b = 2 \qquad 6x_1 + 8x_2 + 2 = 0 \qquad \frac{2}{|W|} = \frac{2}{10}$$



$$Margin = \frac{2}{5} \quad Error = 25$$

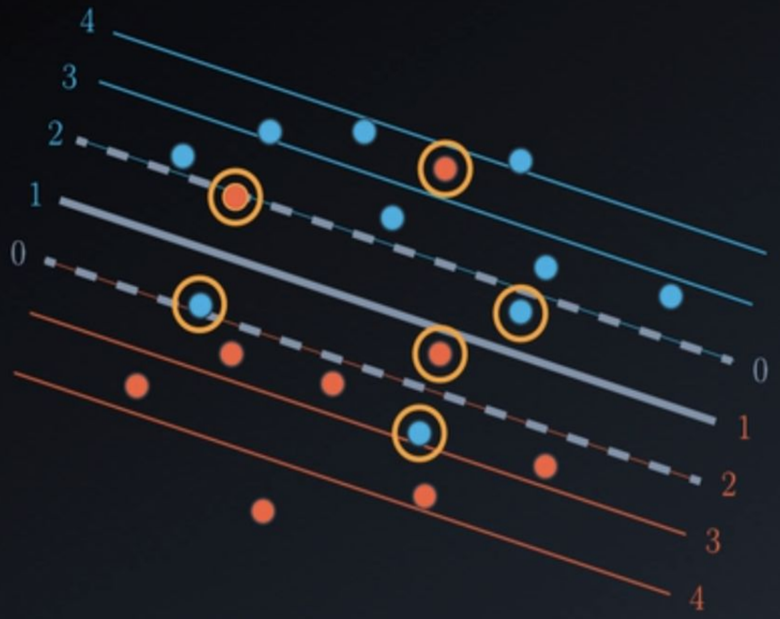
LARGE MARGIN
SMALL ERROR



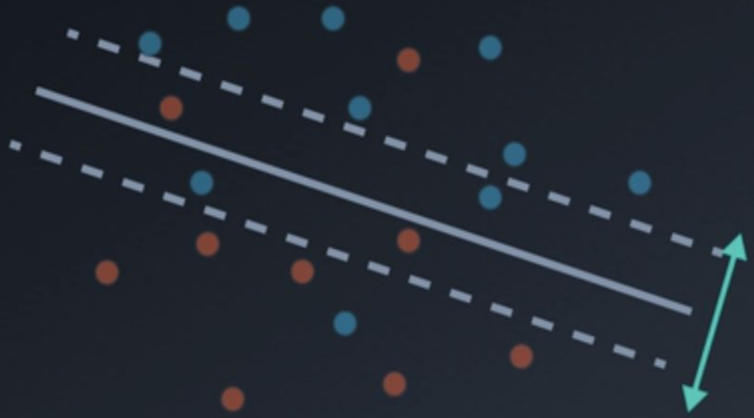
$$Margin = \frac{2}{10} \quad Error = 100$$

SMALL MARGIN
LARGE ERROR

Error Function

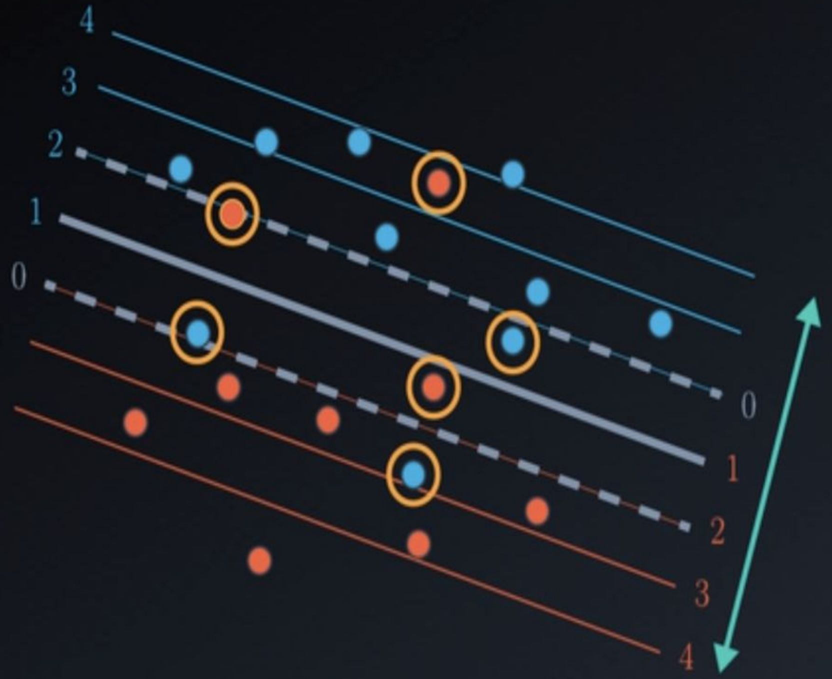


CLASSIFICATION ERROR



MARGIN ERROR

Error Function



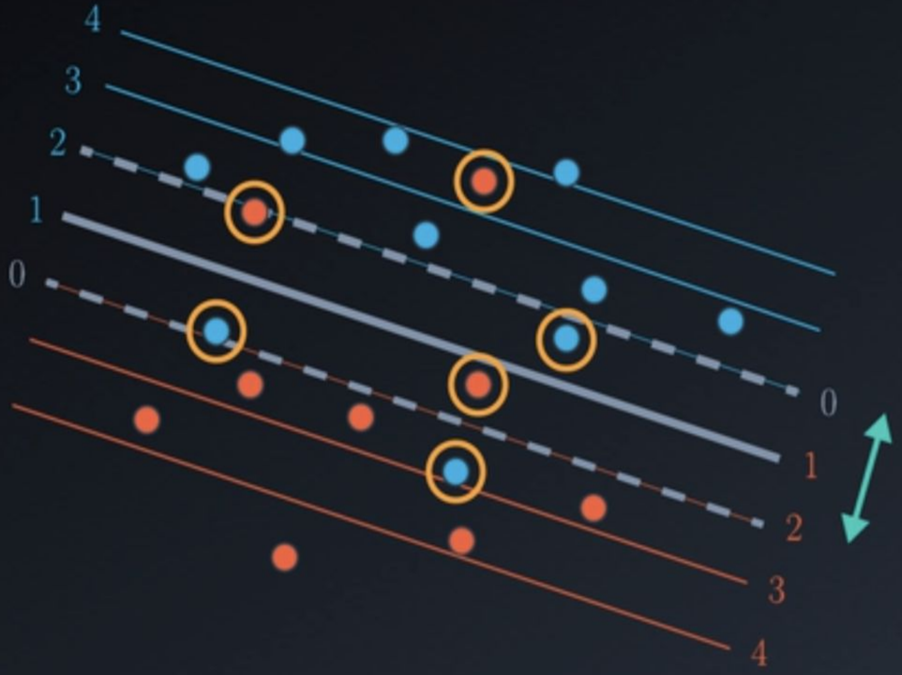
Error = Classification Error + Margin Error

MINIMIZE USING GRADIENT DESCENT

The C Parameter

Quiz: Which one is a better line?

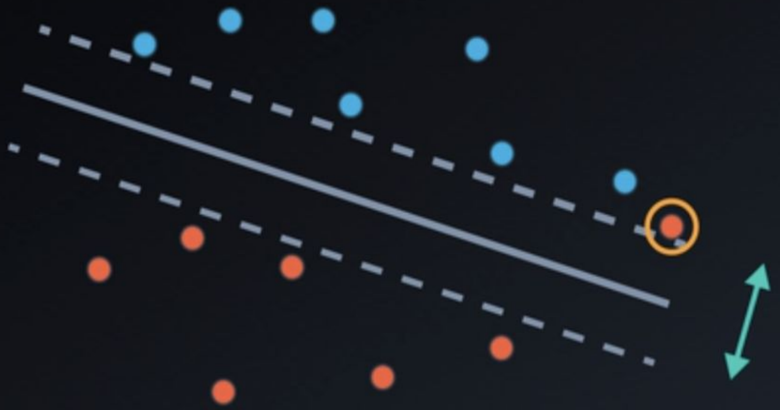
The C Parameter



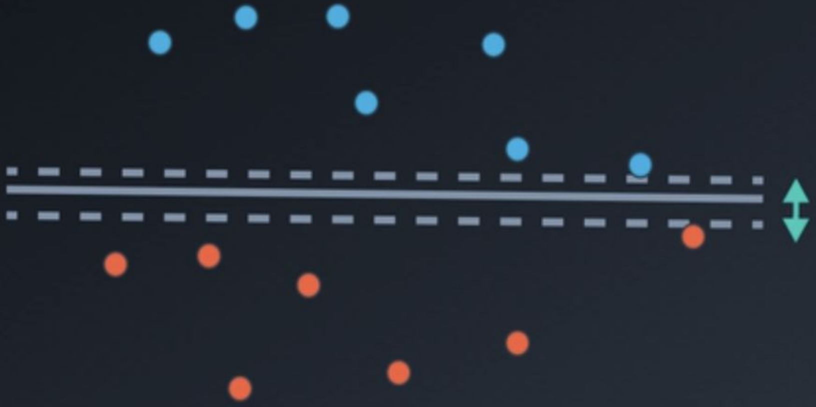
Error = C Classification Error + Margin Error

Large C: Focus on classifying points

The C Parameter



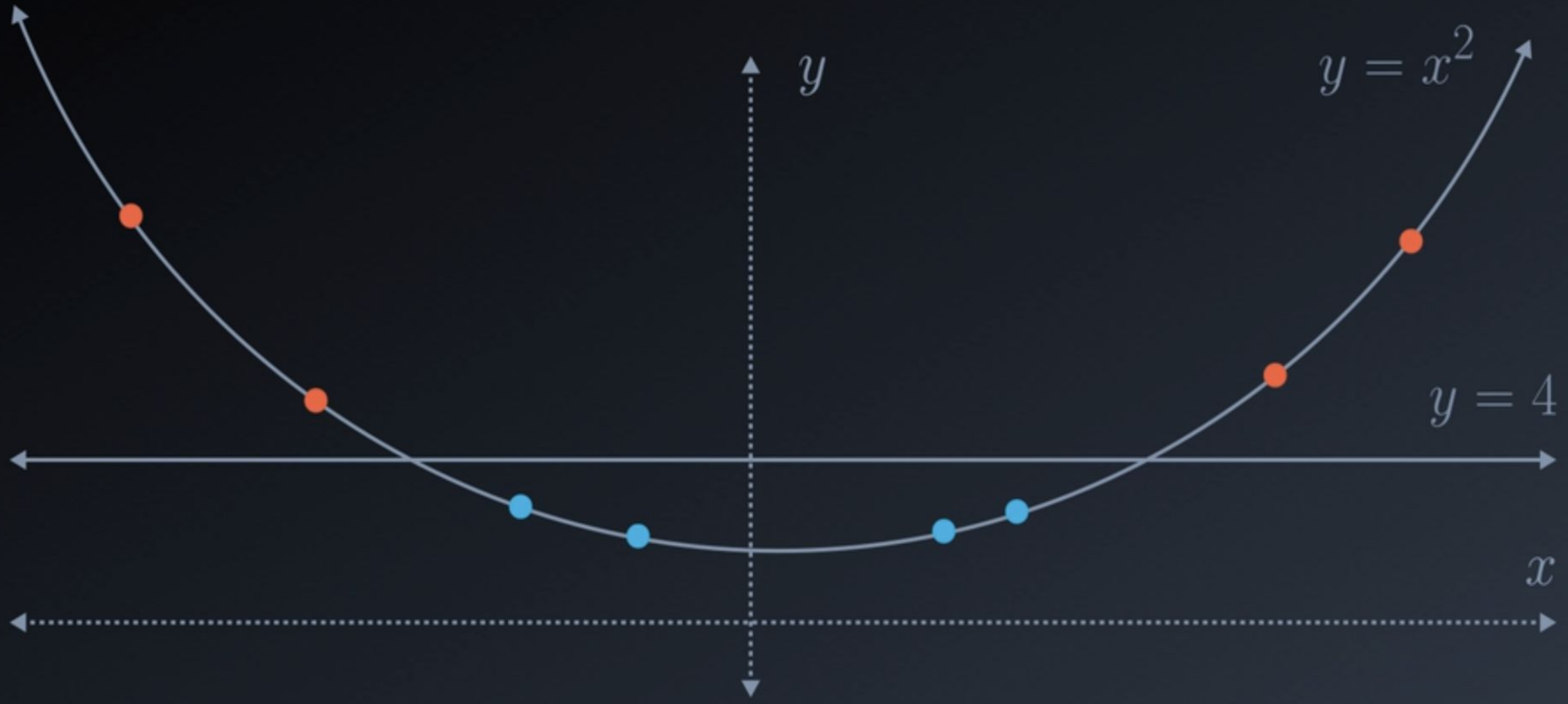
Small C
Large margin
May make classification errors



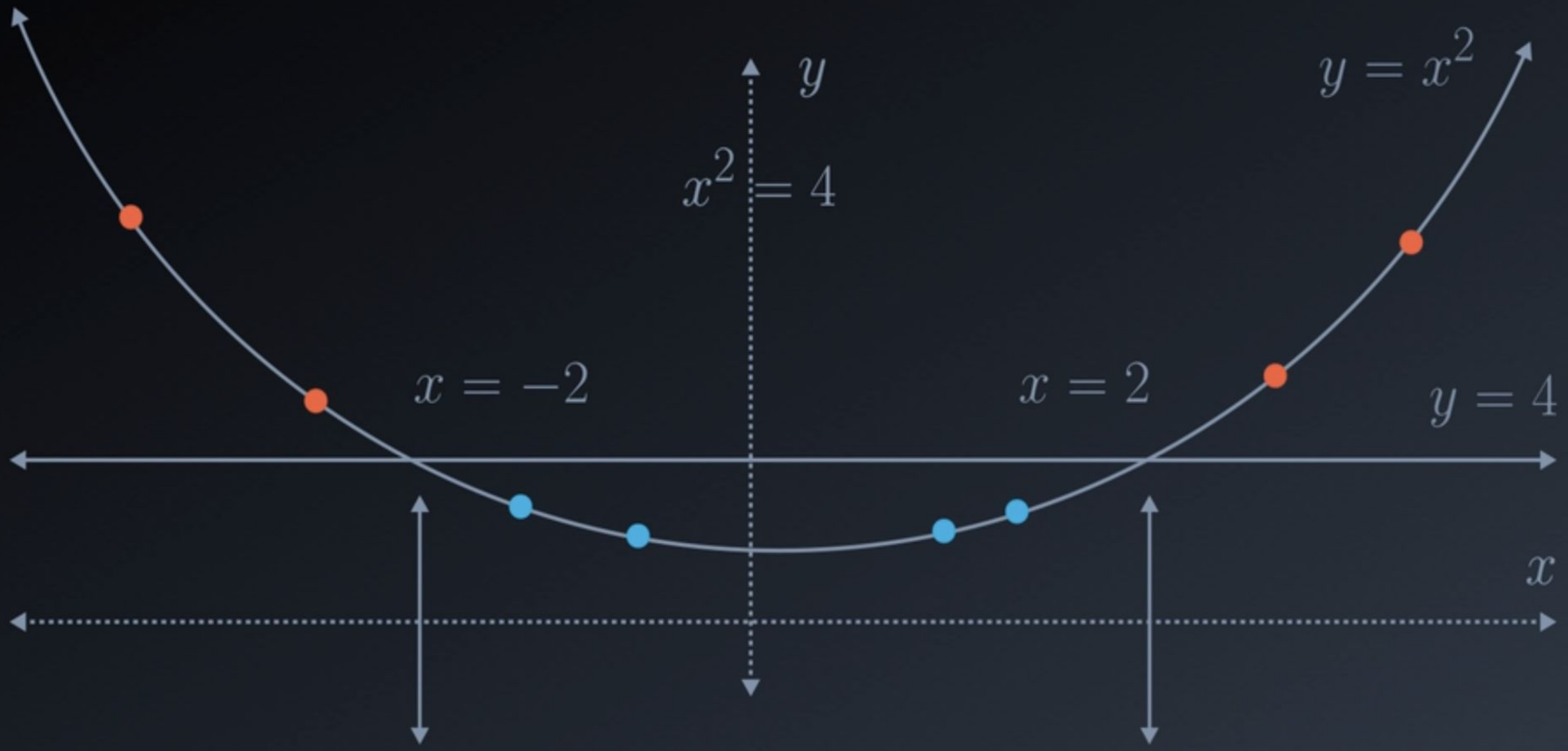
Large C
Classifies points well
May have a small margin

When a line is not enough...

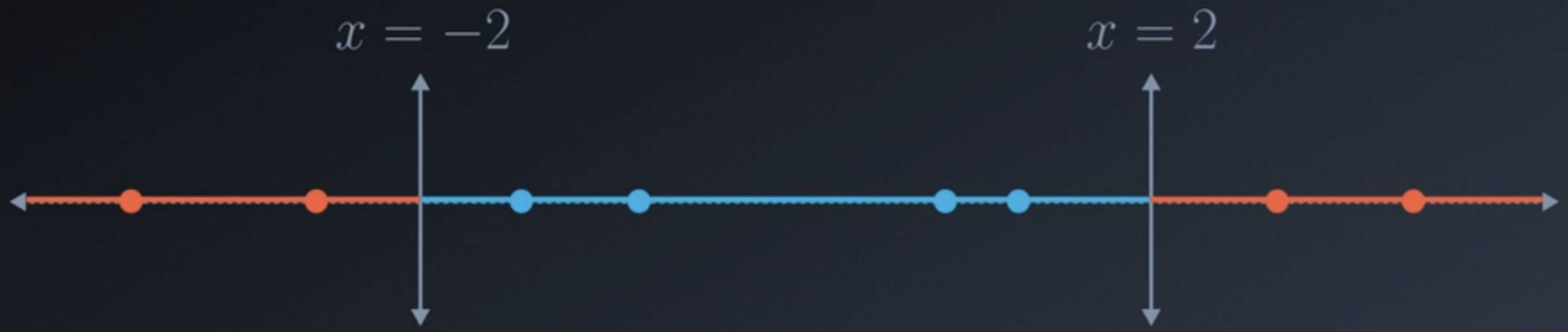
When a line is not enough...



When a line is not enough...

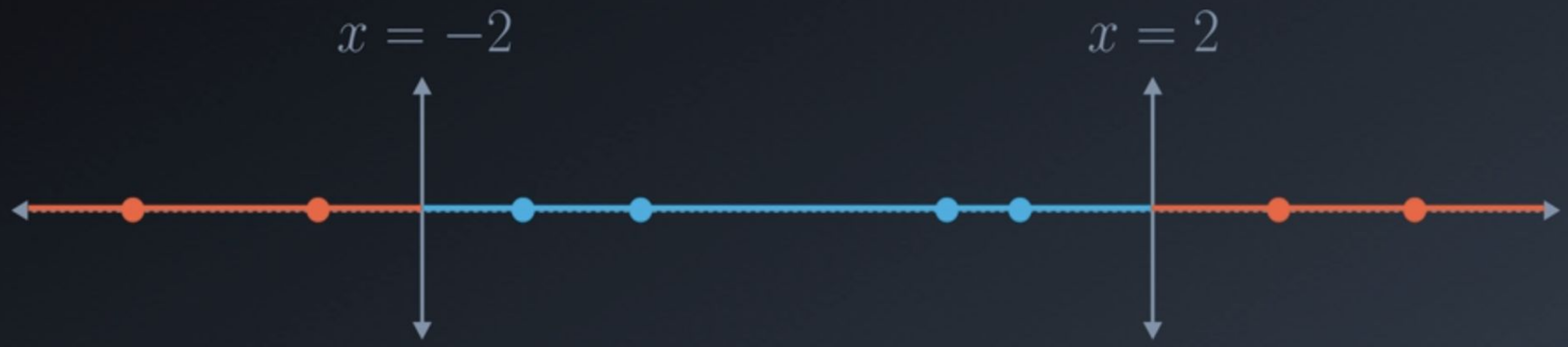


When a line is not enough...

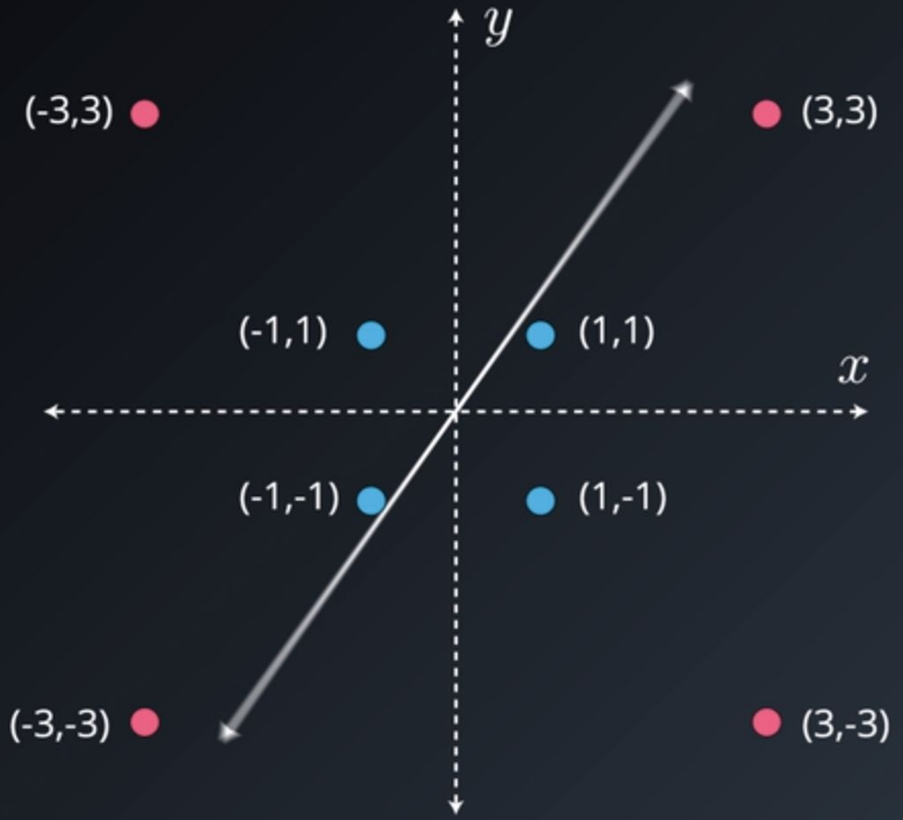


When a line is not enough...

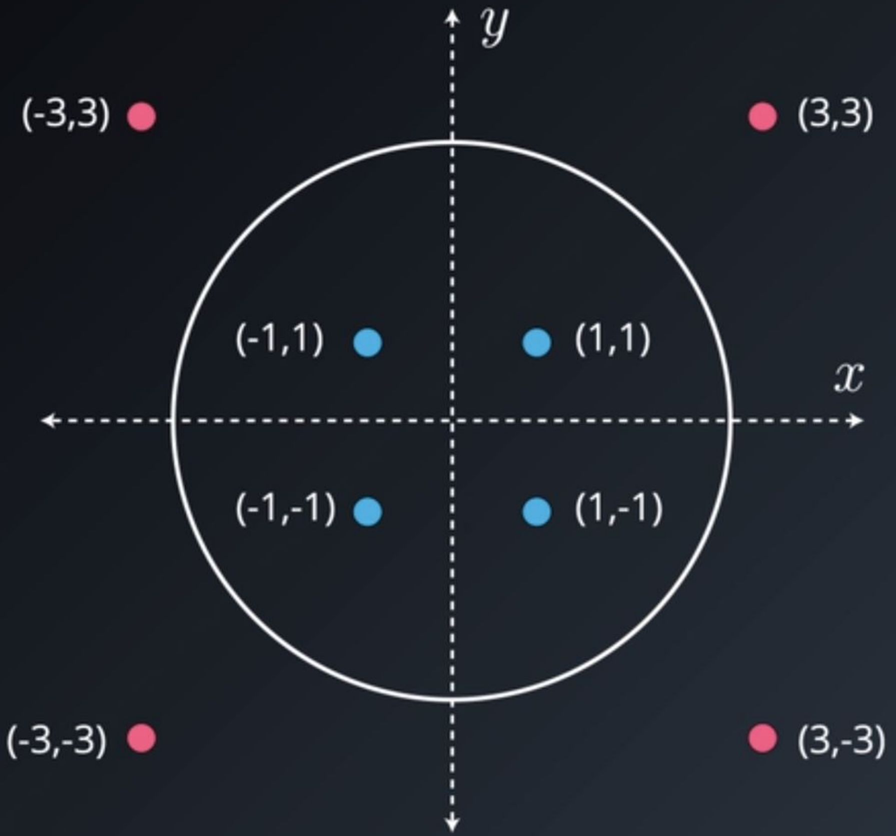
Kernel Trick



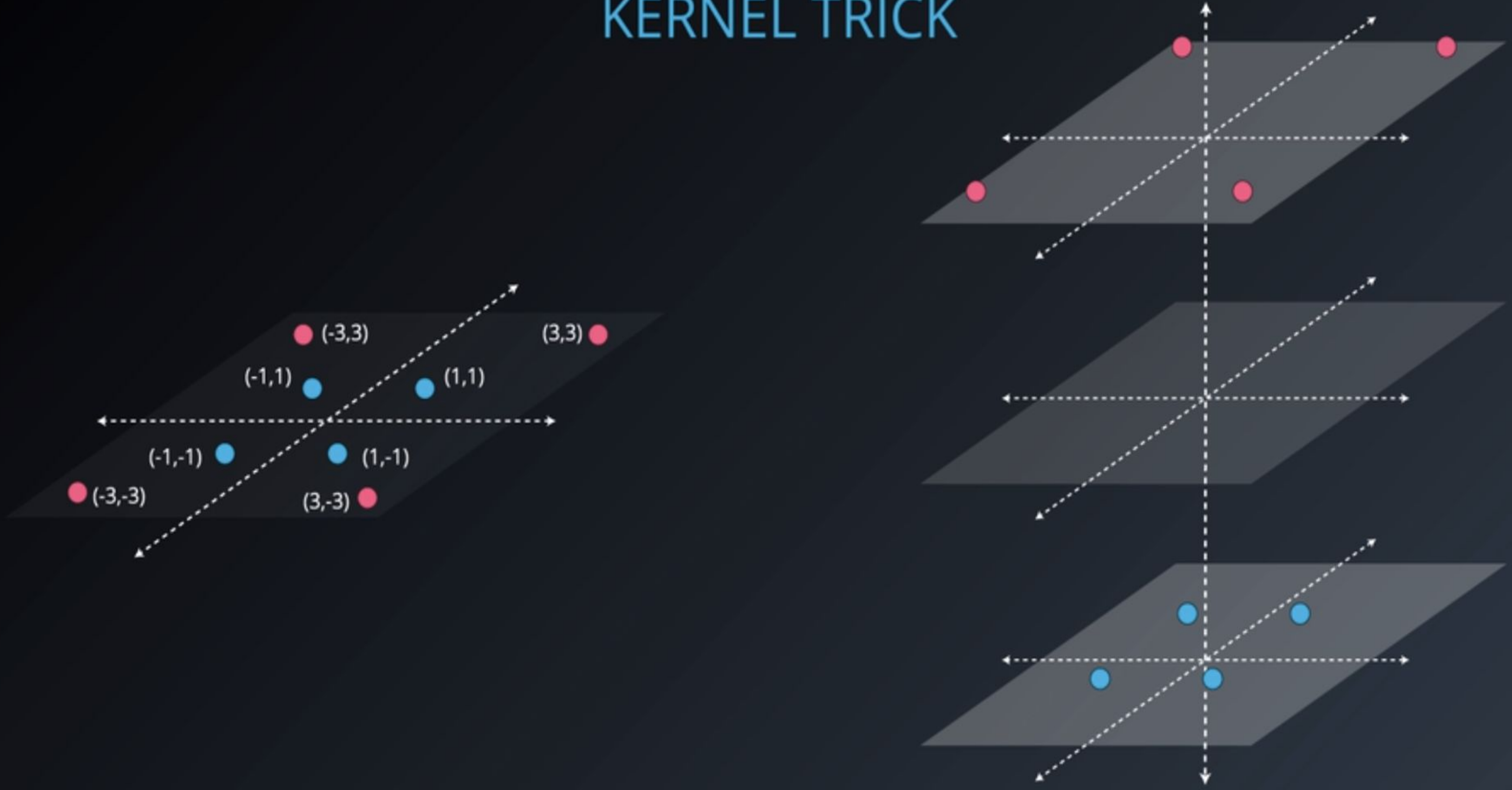
KERNEL TRICK



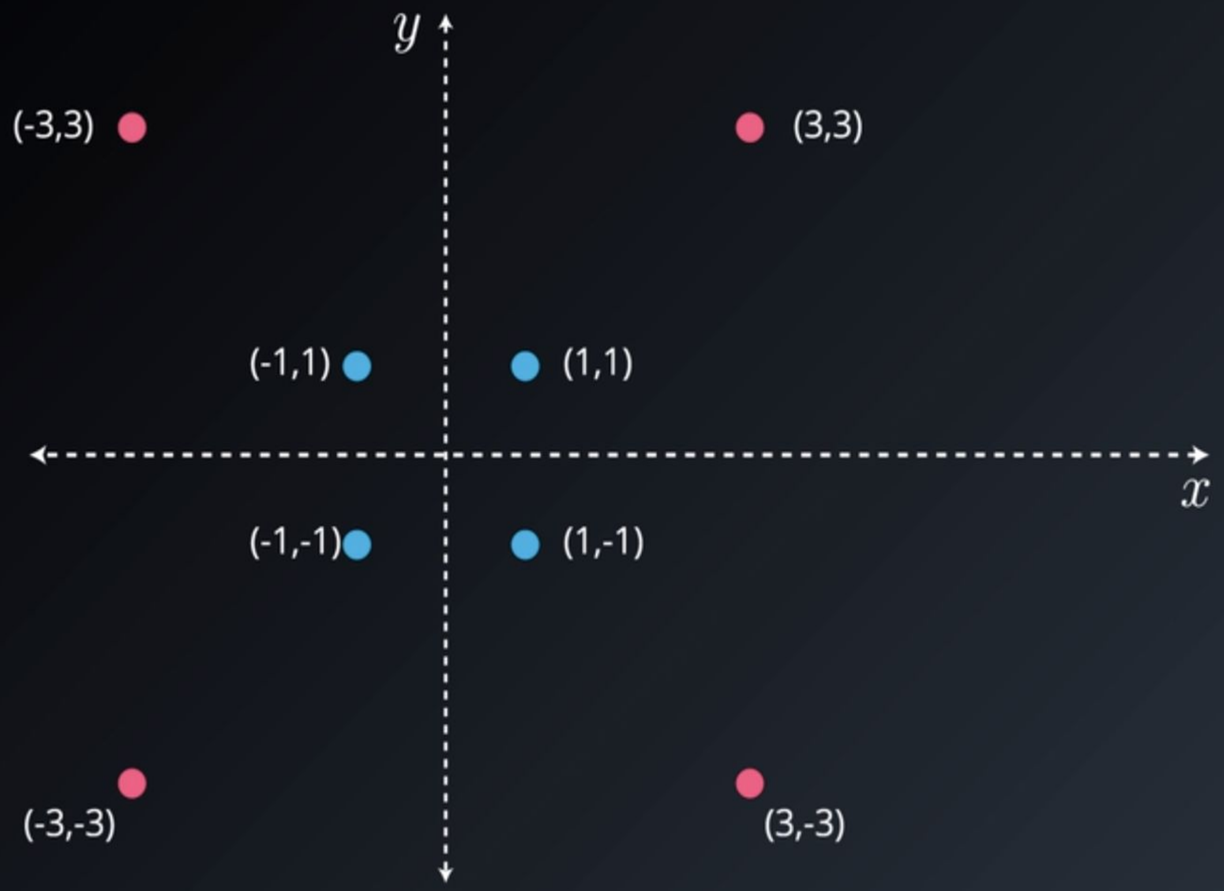
KERNEL TRICK



KERNEL TRICK



KERNEL TRICK



Quiz: Which equation could come to our rescue?

- ○ $x + y$
- ○ xy
- ○ $x^2 + y^2$

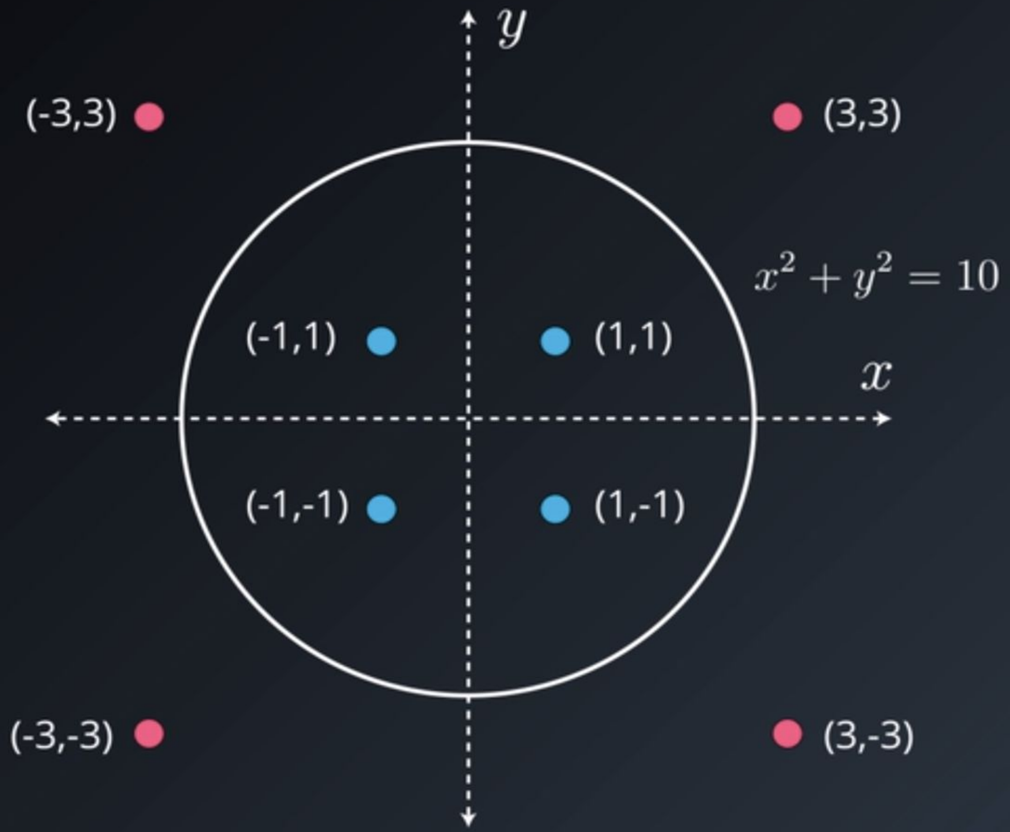
KERNEL TRICK

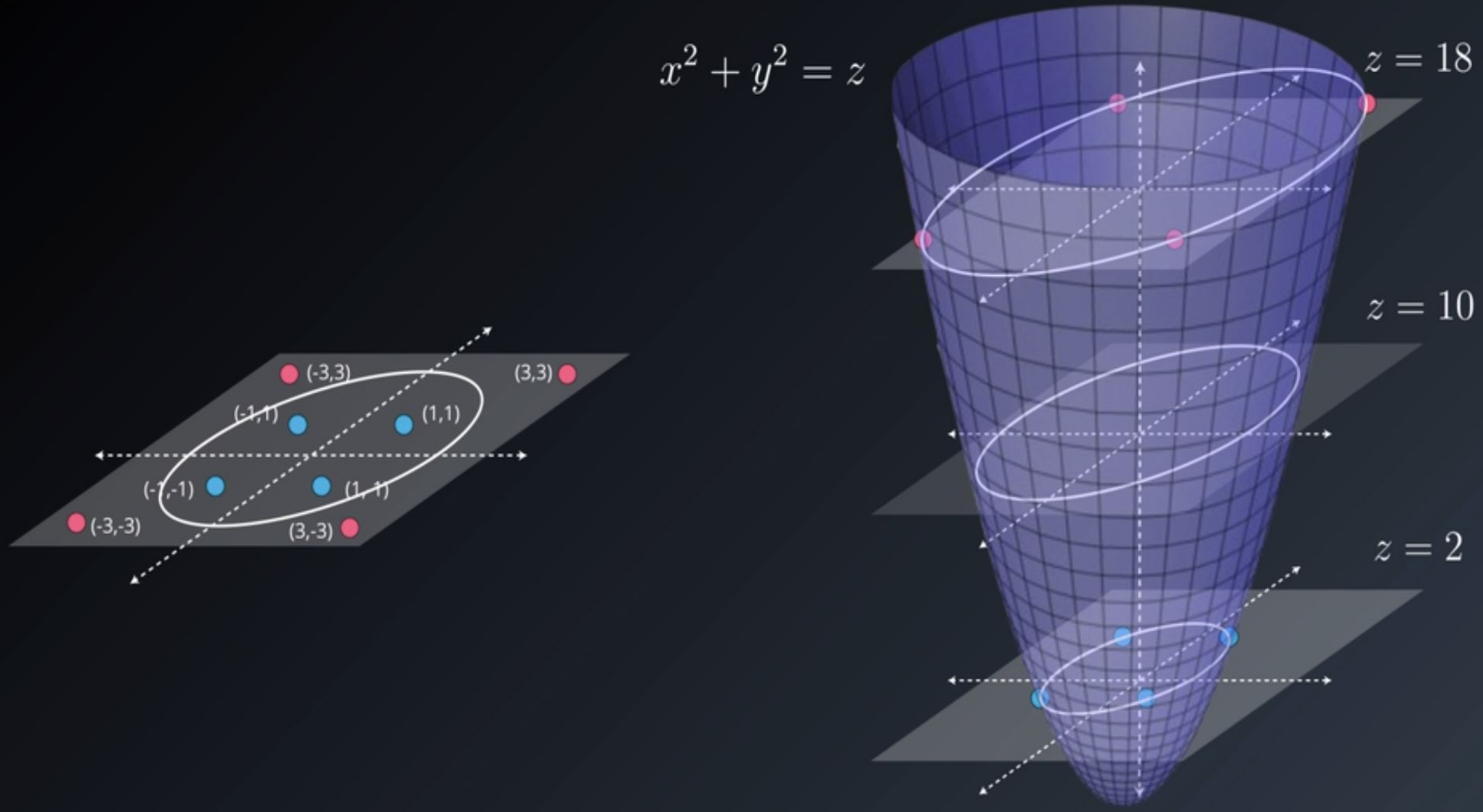
	(-3,-3)	(3,-3)	(-3,3)	(3,3)	(-1,-1)	(1,-1)	(-1,1)	(1,1)
$x+y$	-6	0	0	6	-2	0	0	2
xy	9	-9	-9	9	1	-1	-1	1
x^2+y^2	18	18	18	18	2	2	2	2

Quiz: Which equation could come to our rescue?

- ○ $x+y$
- ○ xy
- ☑ x^2+y^2

KERNEL TRICK



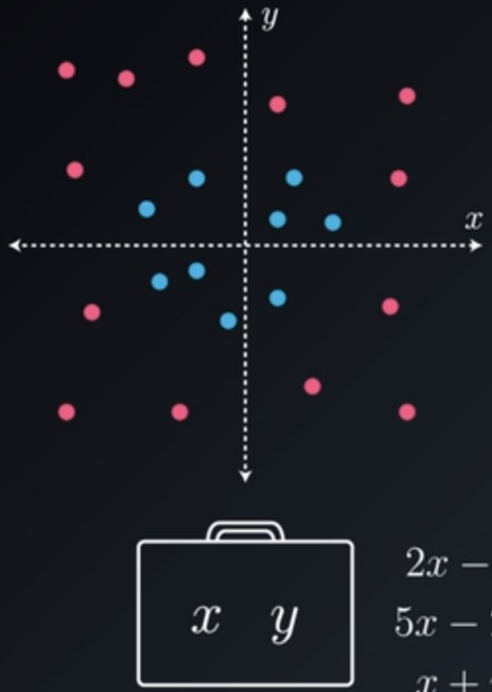


(-3,3)
(3,3)
(-1,1)
(1,1)
(-1,-1)
(1,-1)
(-3,-3)
(3,-3)
$x^2 + y^2 = z$
$z = 18$
$z = 10$
$z = 2$

KERNEL TRICK

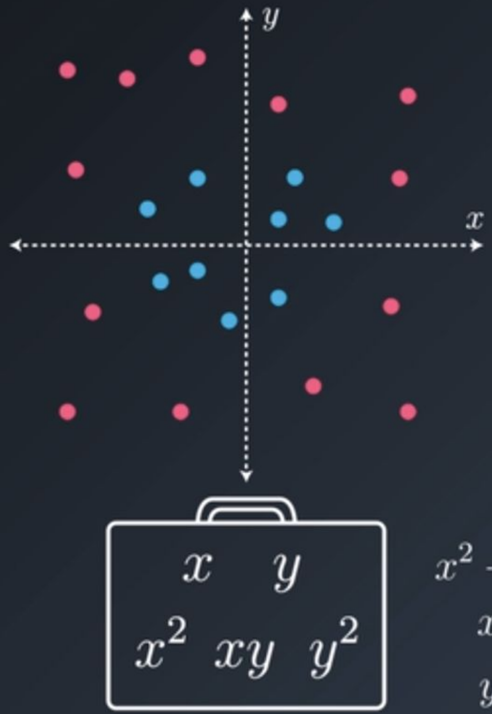
2 Dimensions		5 Dimensions
(x, y)	$\longrightarrow$	(x, y, x^2, xy, y^2)
(2, 3)		(2, 3, 4, 6, 9)
Degree 2 Polynomial boundary	$\longleftarrow$	4-dimensional boundary hyperplane

KERNEL



$$2x - y = 1$$
$$5x - 2y = 3$$
$$x + y = 4$$

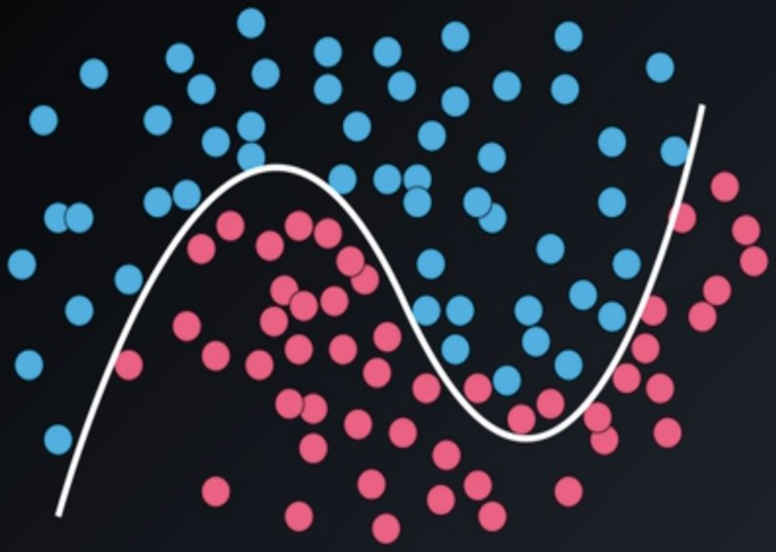
Linear Kernel



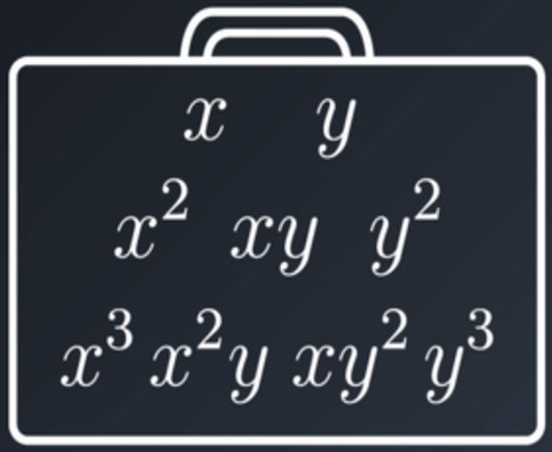
$$x^2 + y^2 = 1$$
$$xy = 1$$
$$y = x^2$$

Polynomial Kernel

KERNEL TRICK

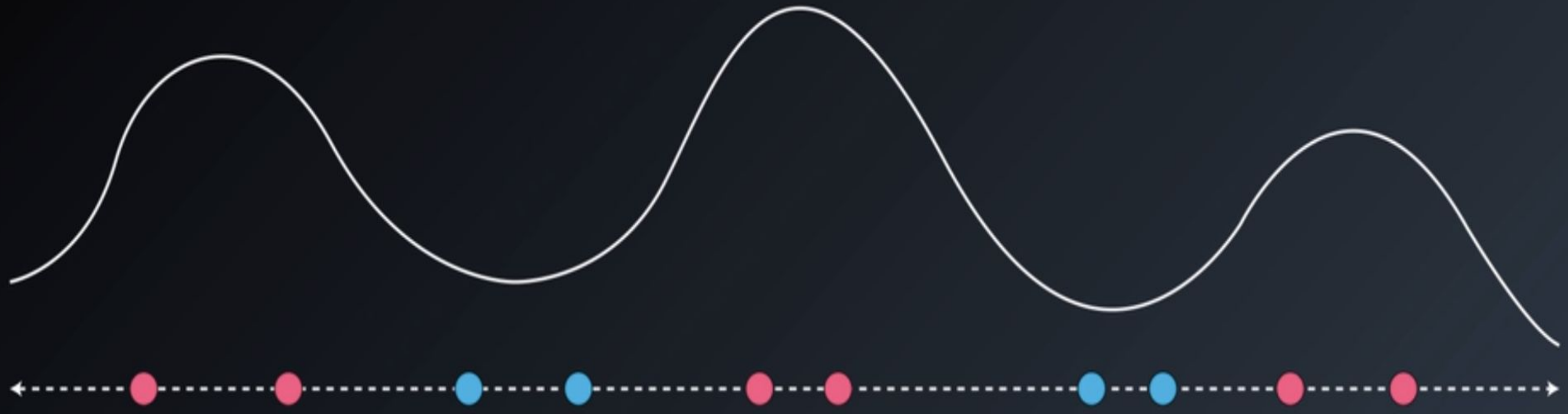


$$y = x^3 + 2x^2 - x - 2$$



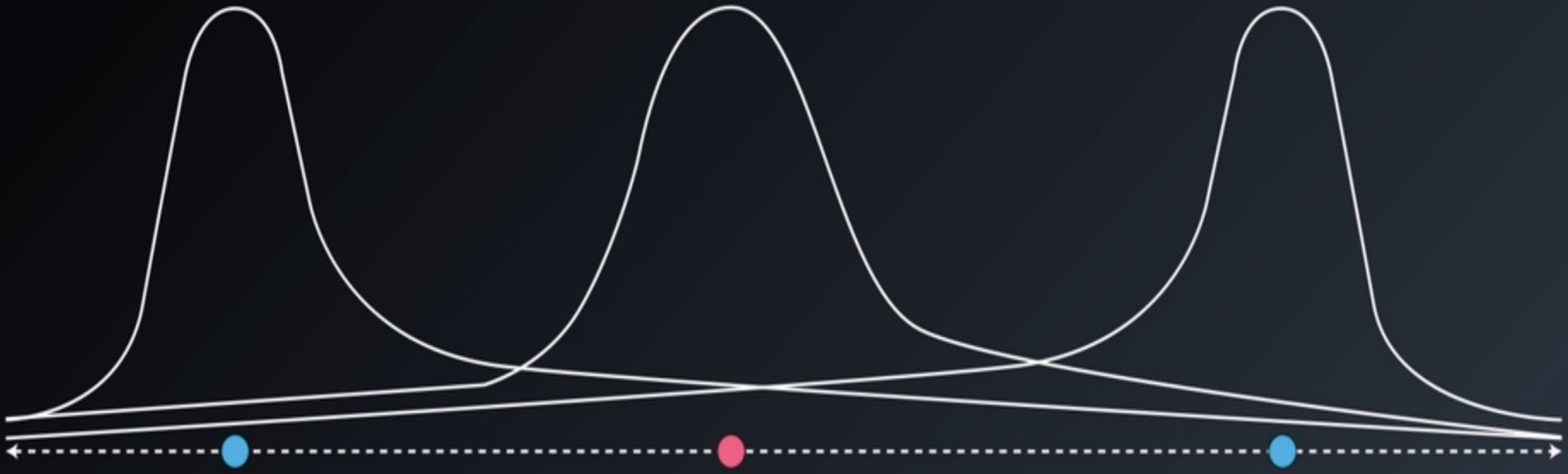
Polynomial Kernel (degree 3)

RBF KERNEL

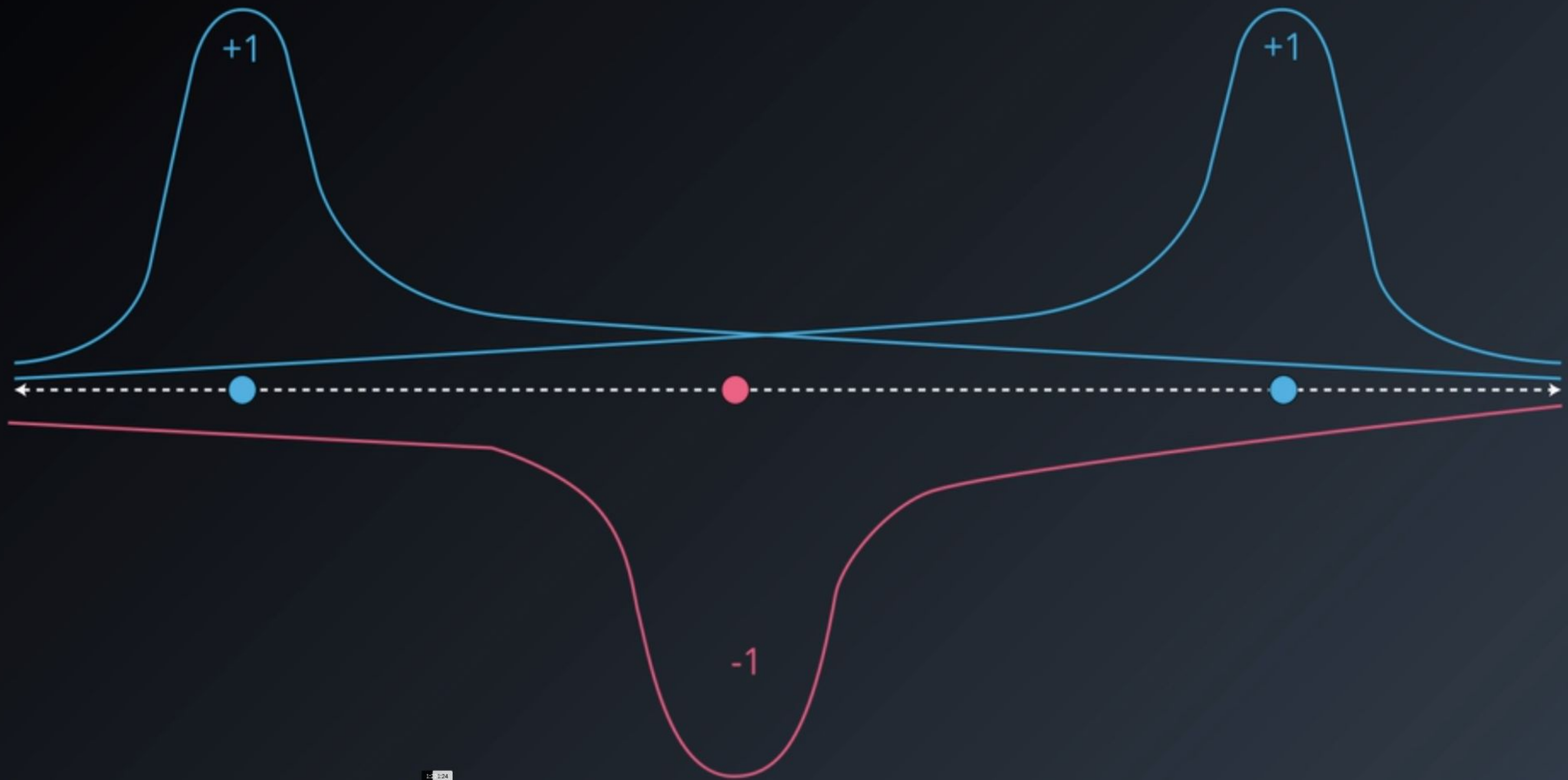


RBF KERNEL

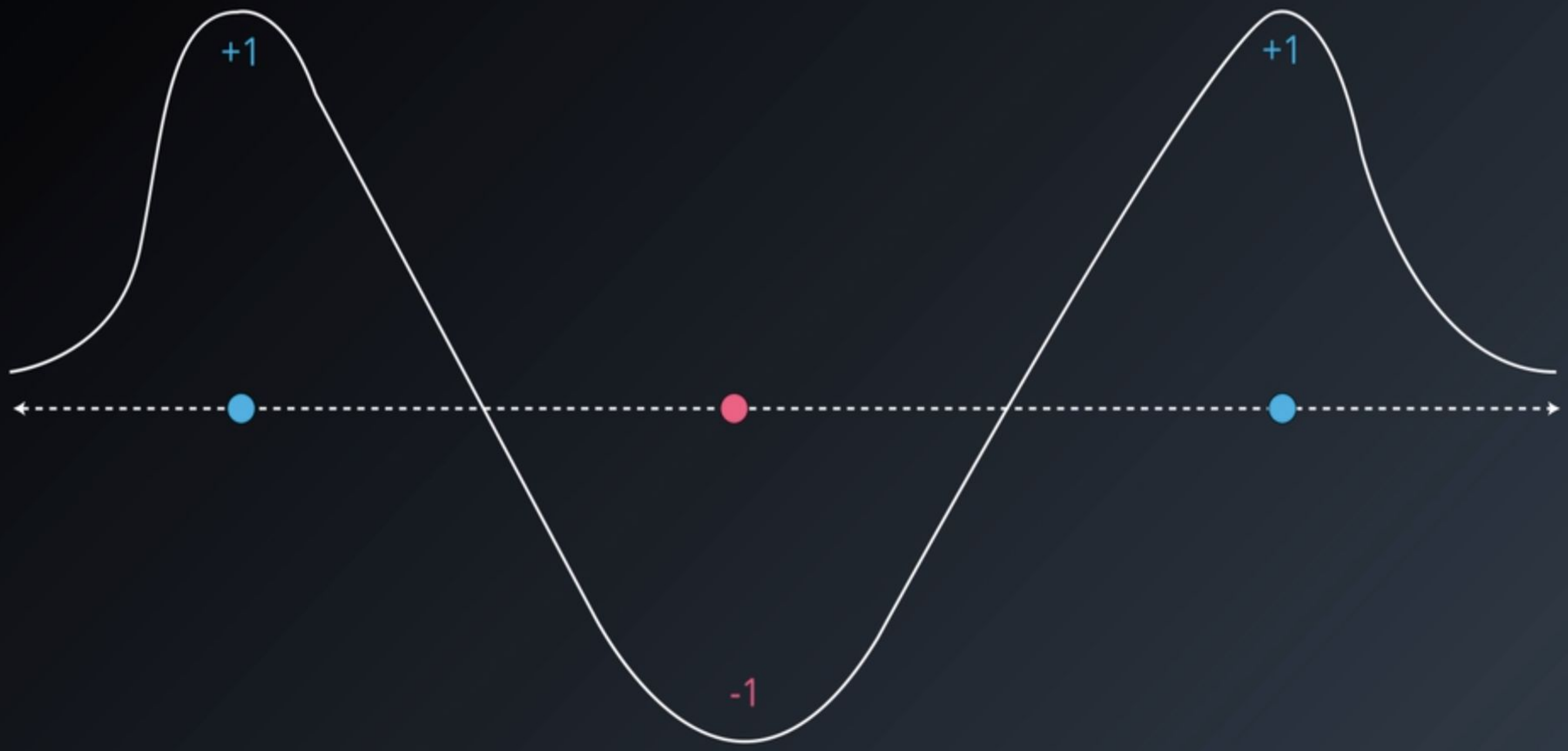
RBF KERNEL



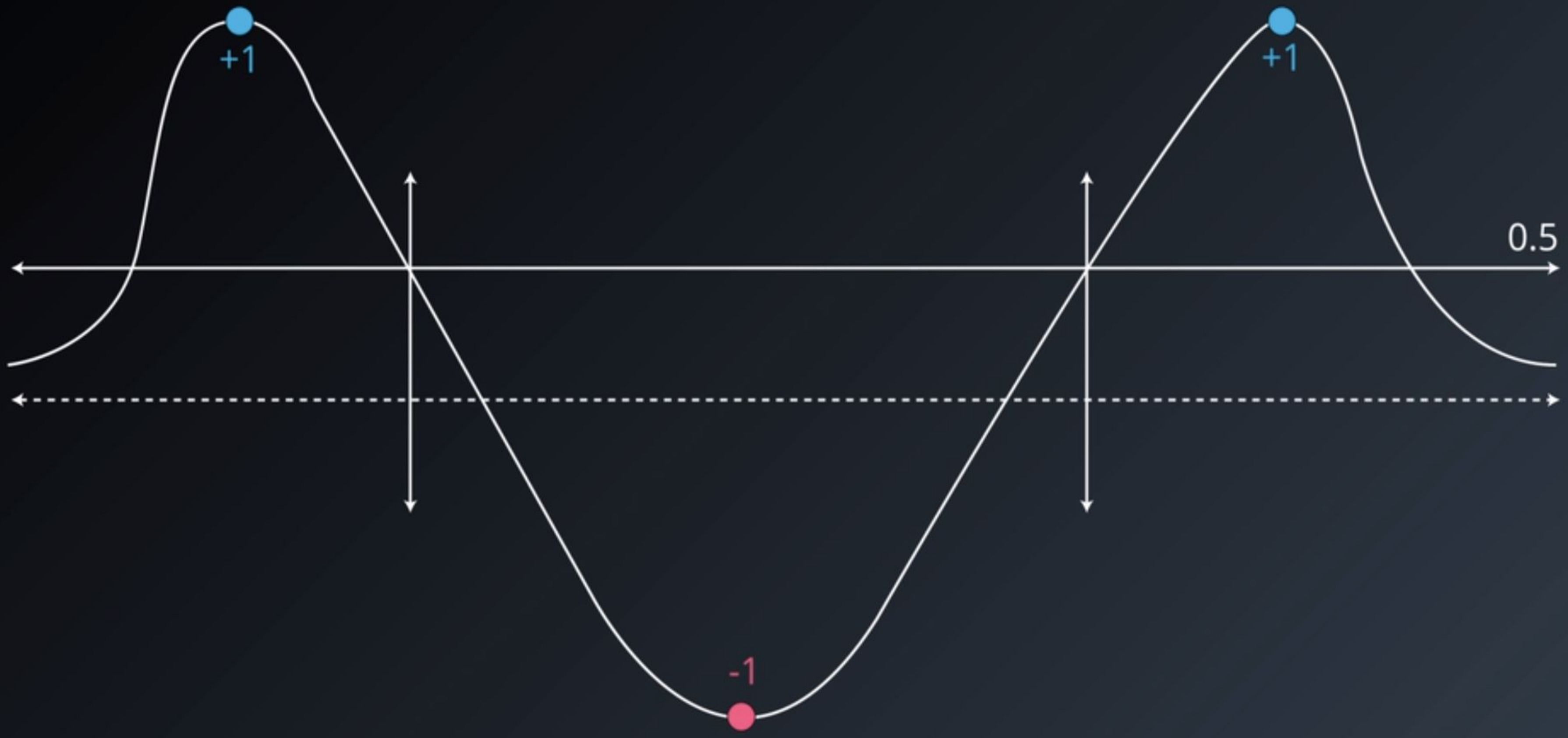
RBF KERNEL



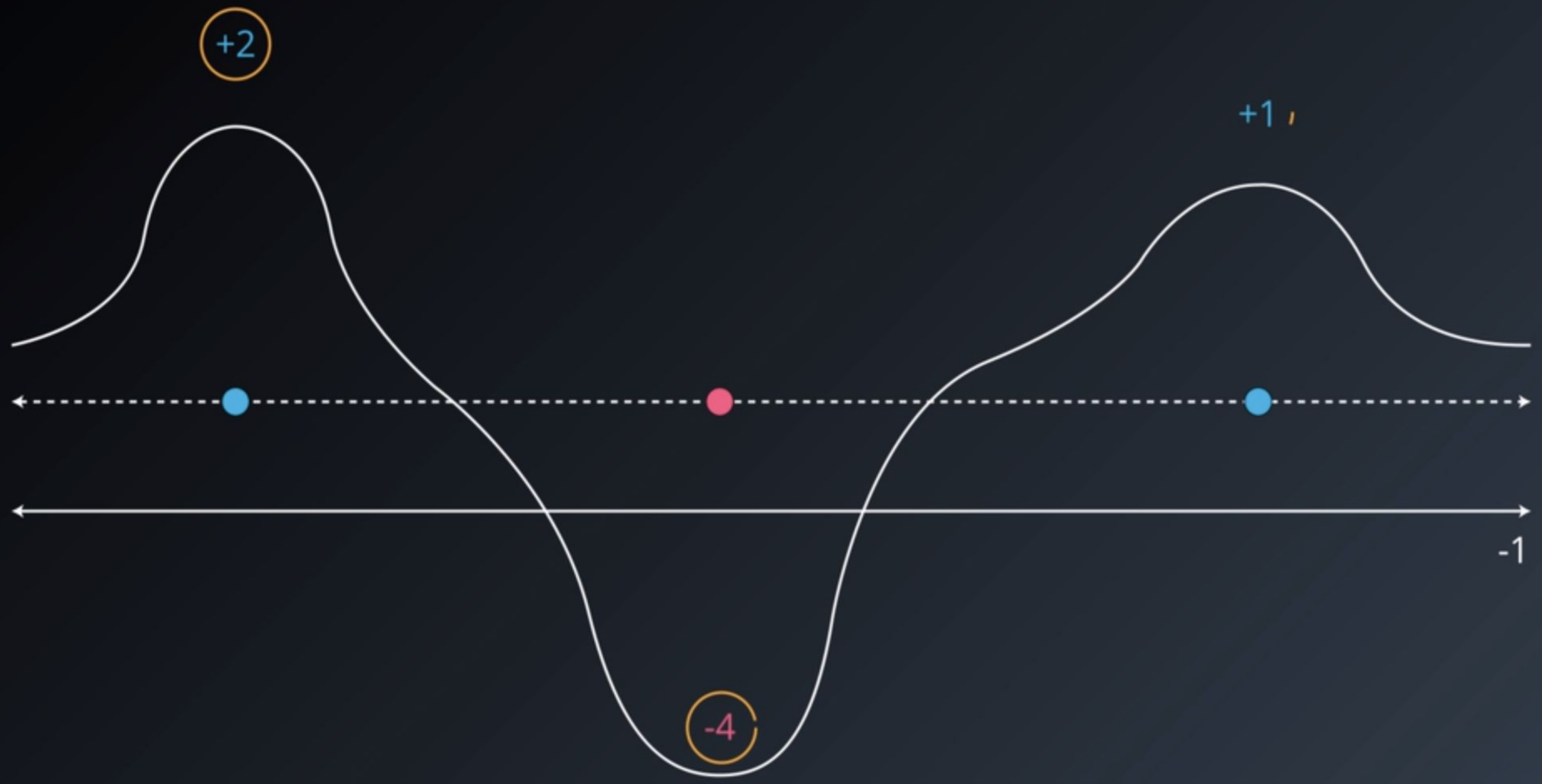
RBF KERNEL



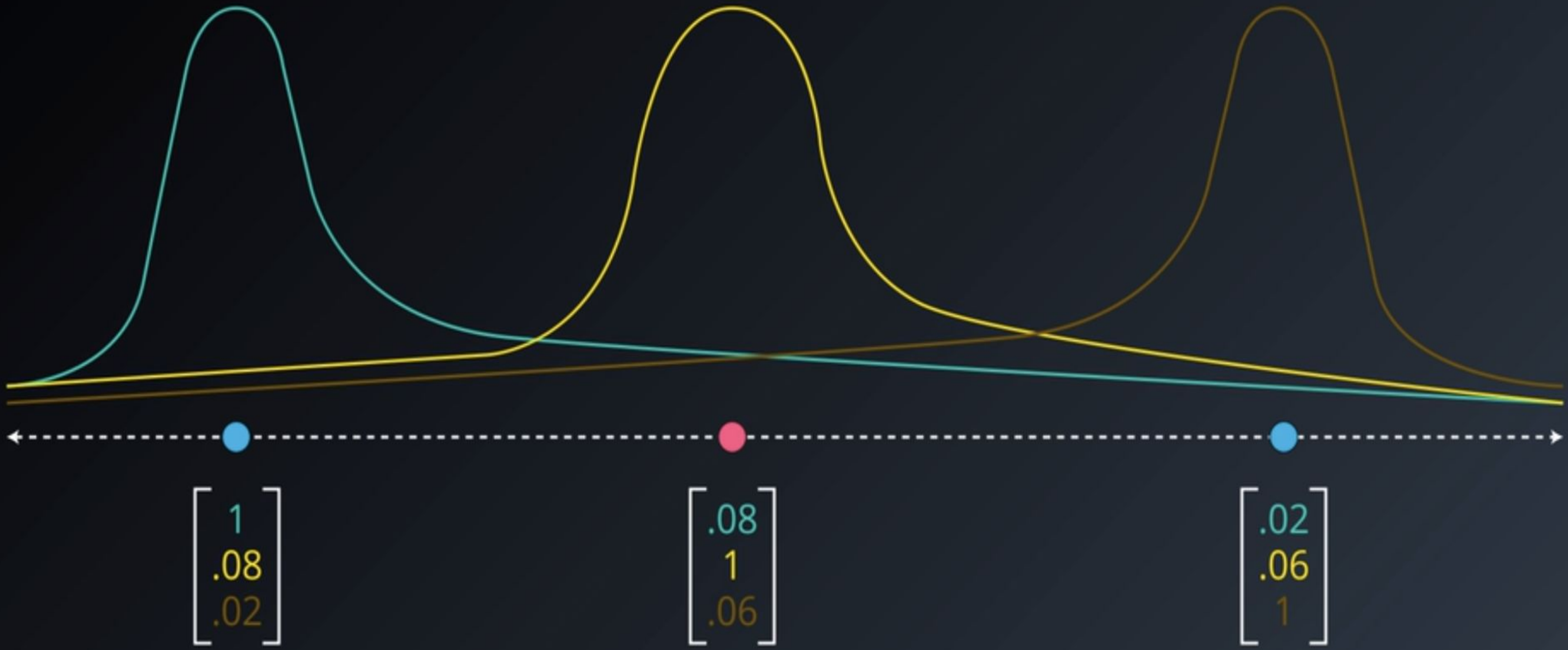
RBF KERNEL



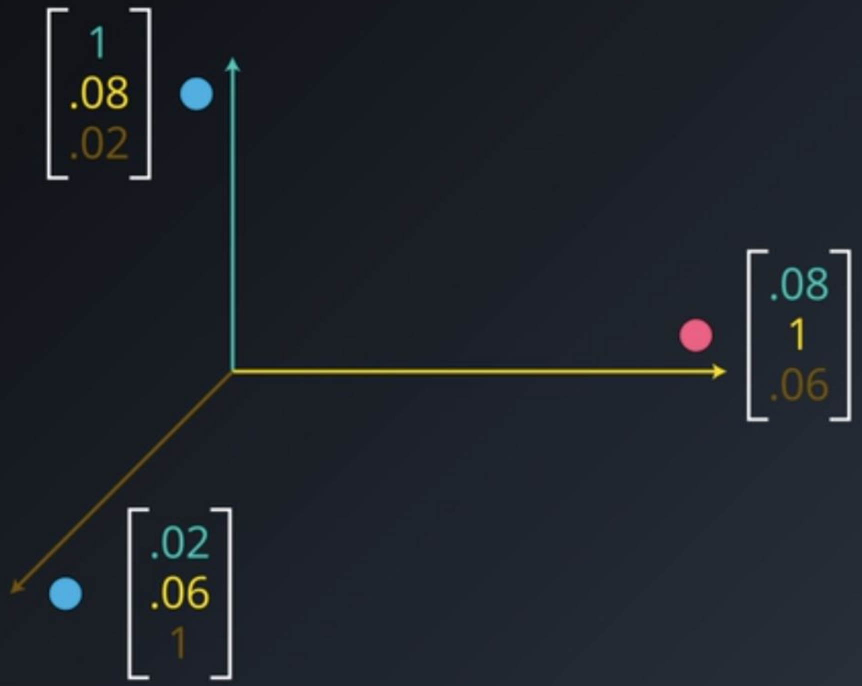
RBF KERNEL



RBF KERNEL

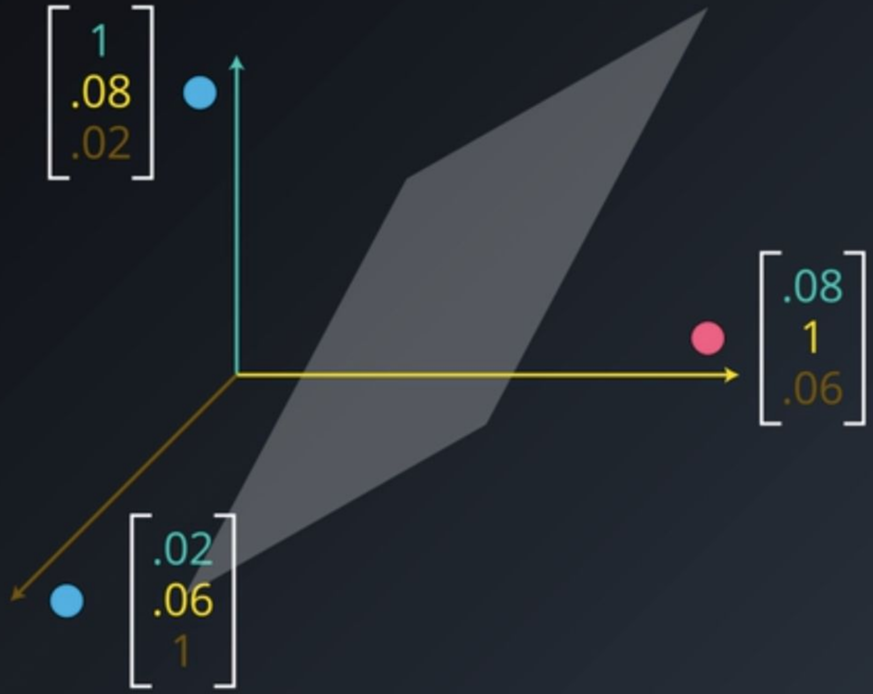


RBF KERNEL



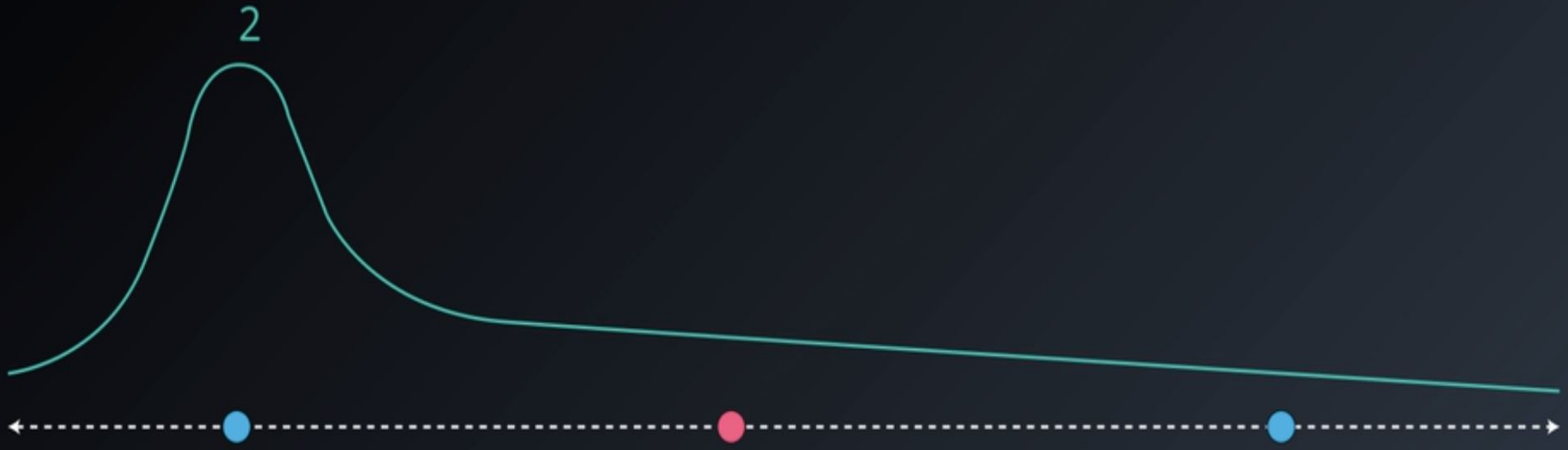
RBF KERNEL

$2x - 4y + 1z = -1$



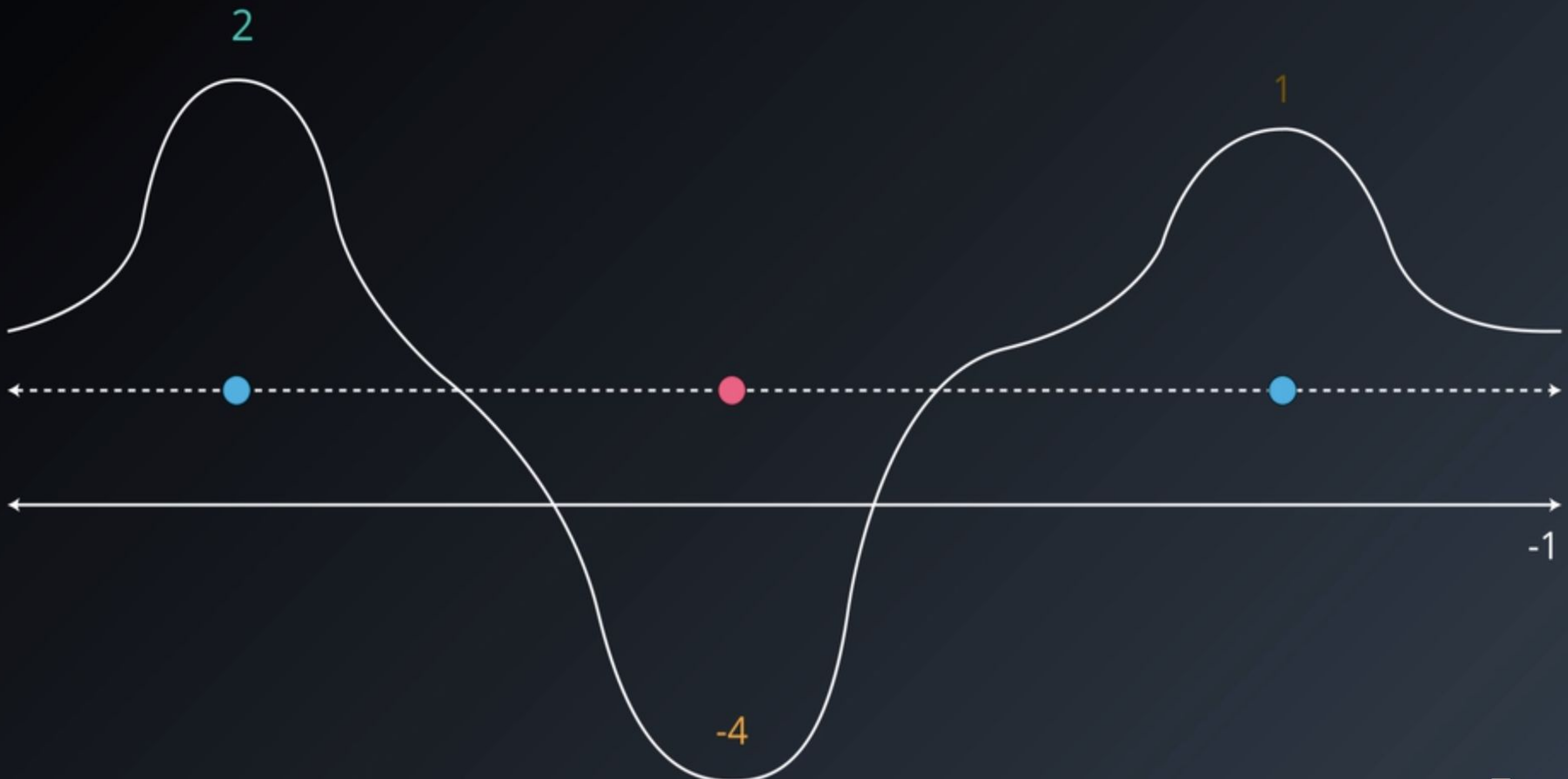
RBF KERNEL

$2x - 4y + 1z = -1$

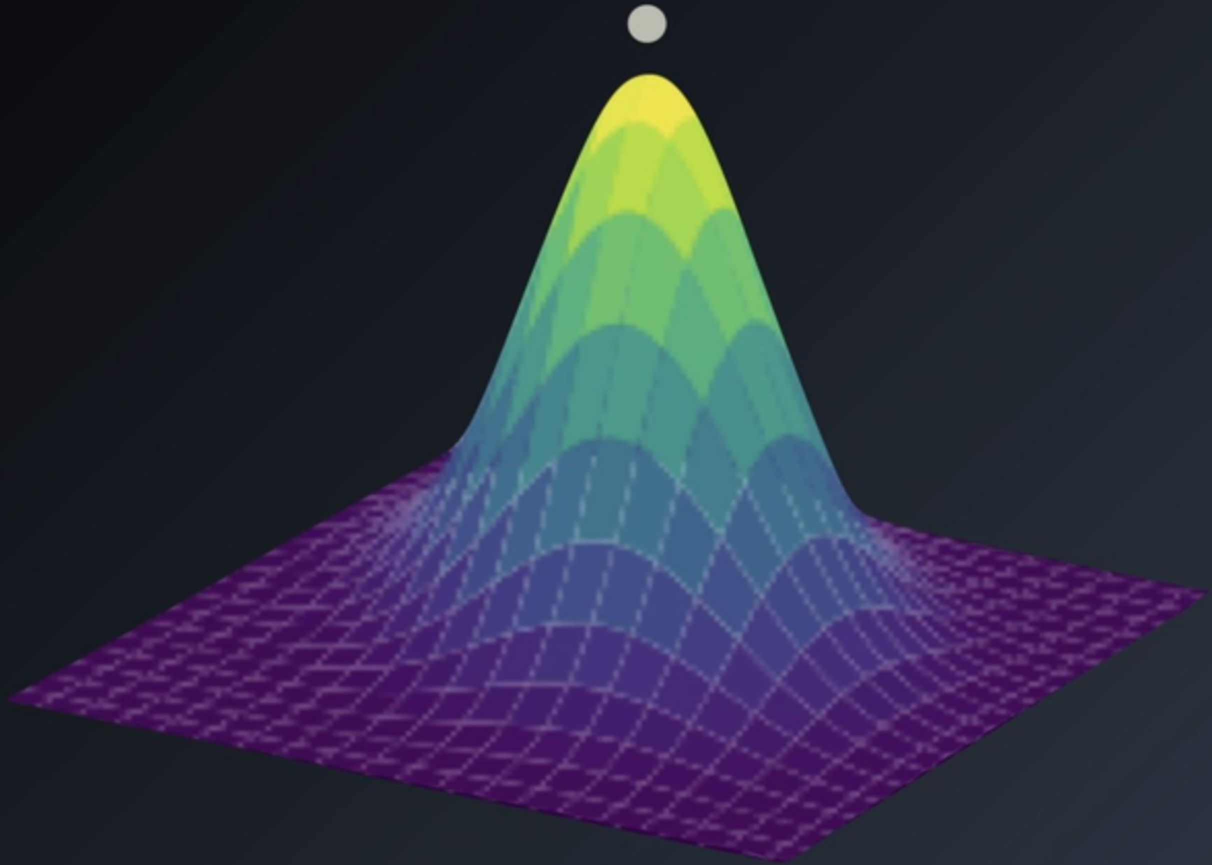


RBF KERNEL

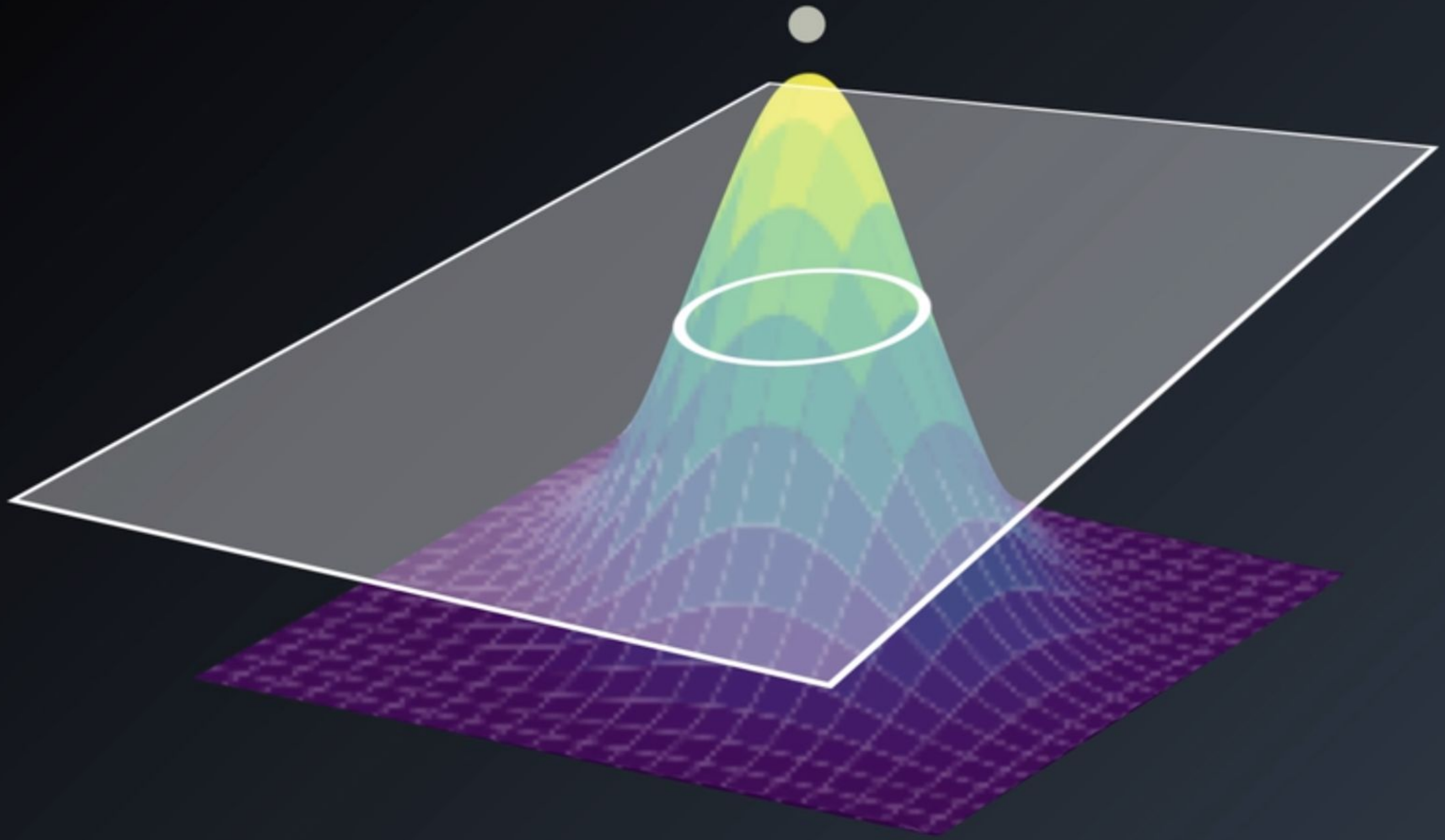
$2x - 4y + 1z = -1$



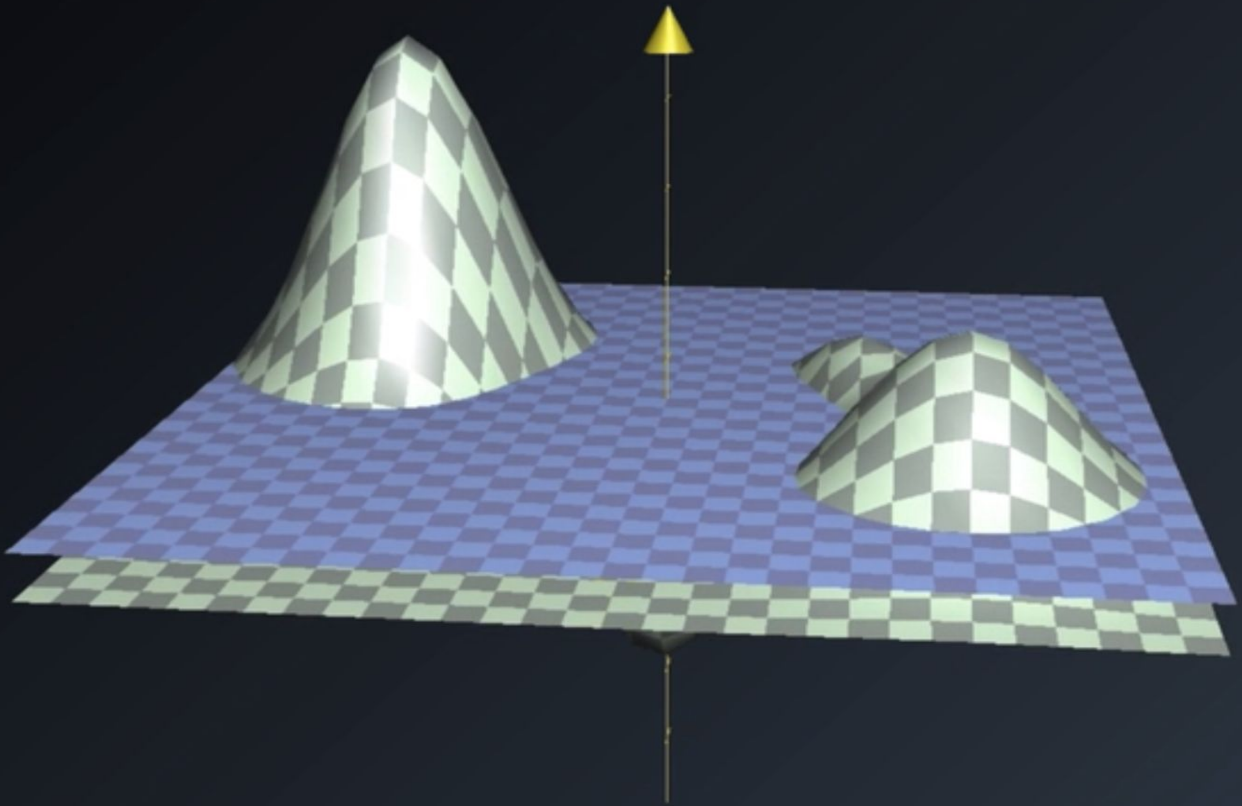
RBF KERNEL



RBF KERNEL

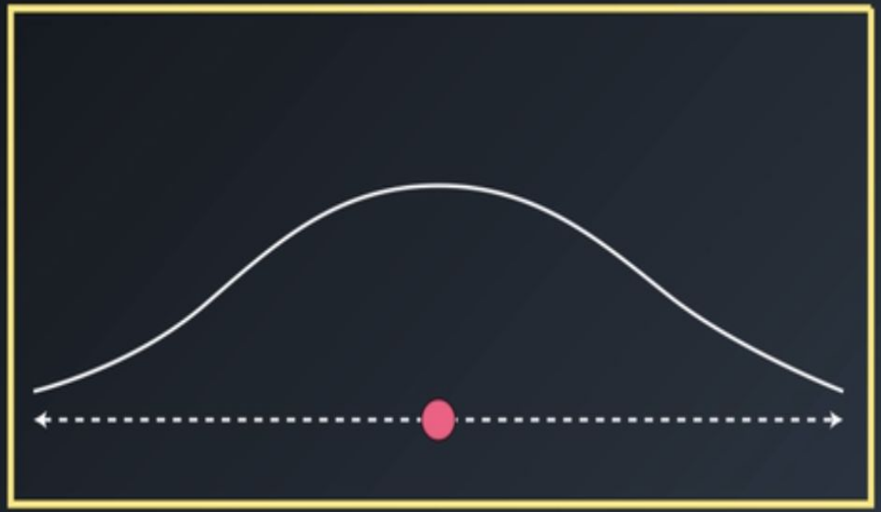


RBF KERNEL



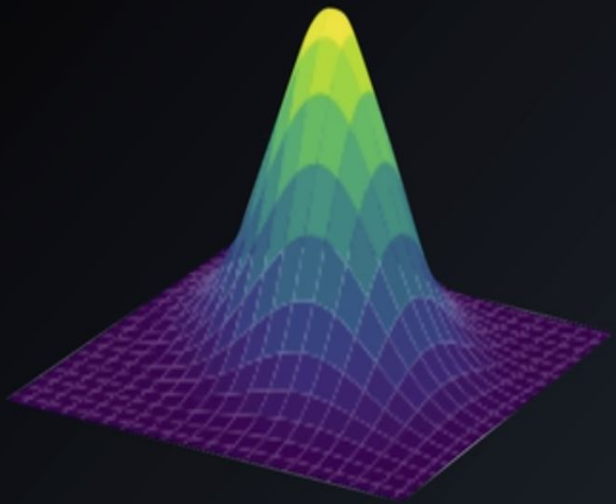
γ PARAMETER

Large γ

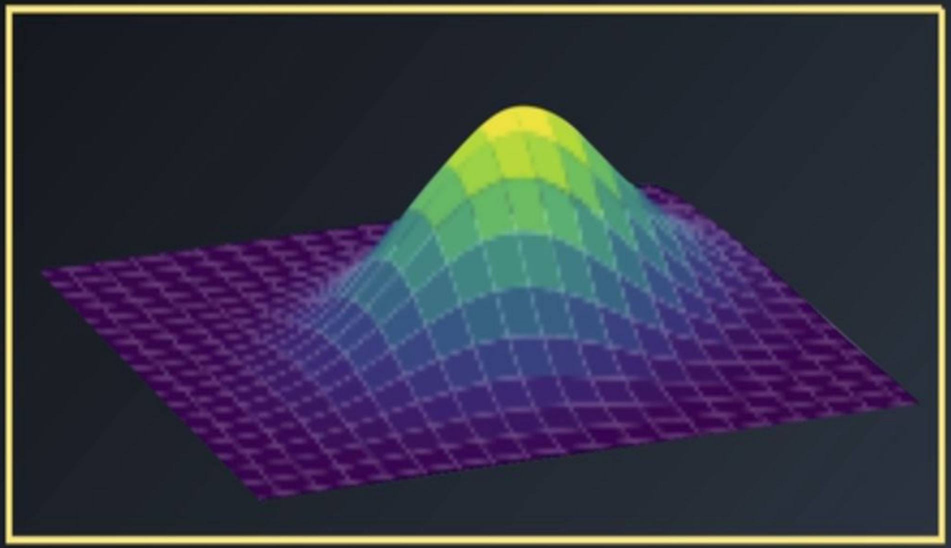


Small γ

RBF KERNEL

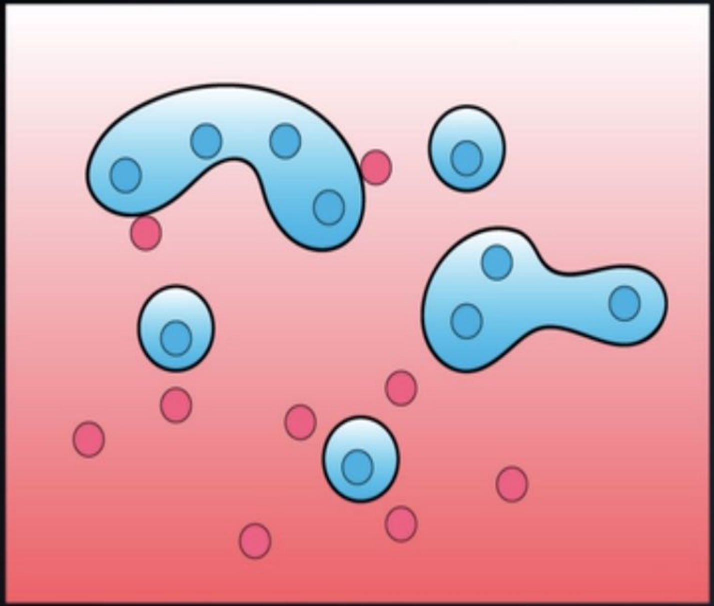


Large γ

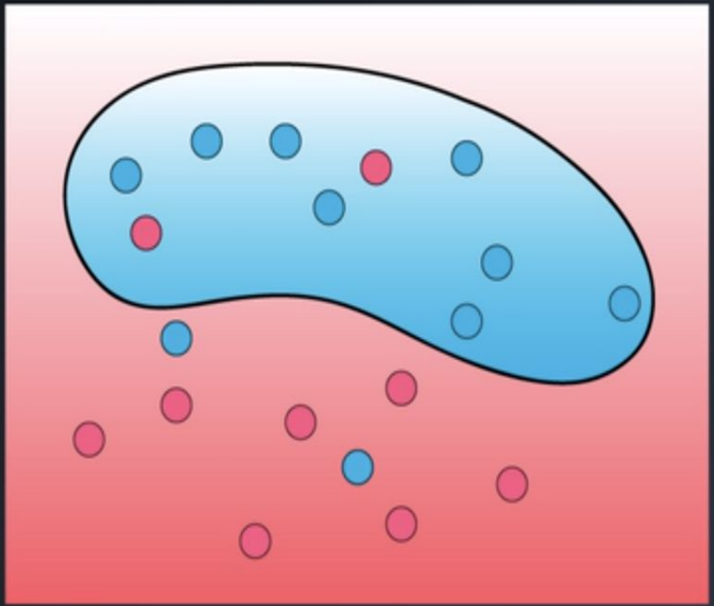


Small γ

RBF KERNEL



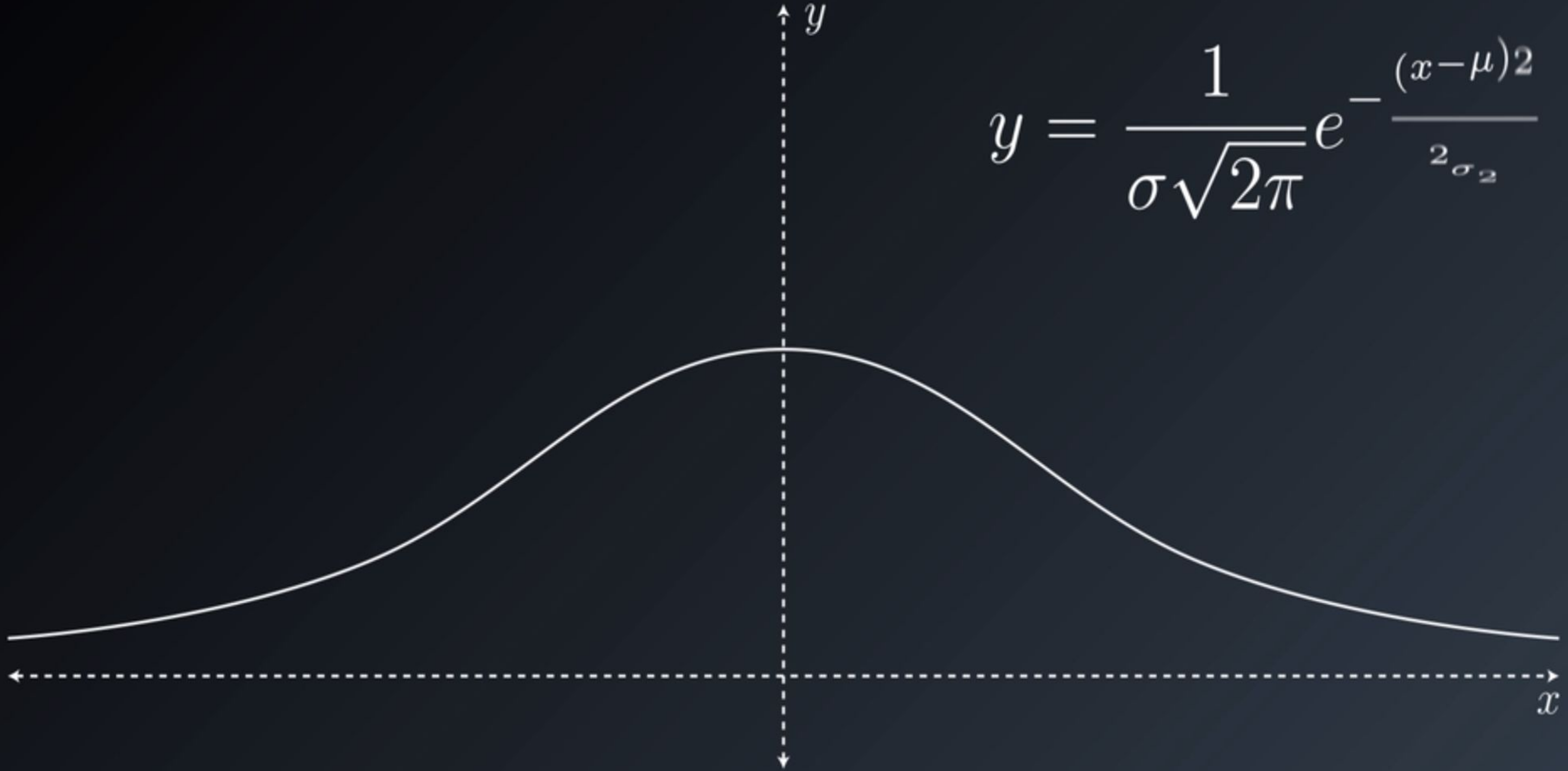
Large γ



Small γ

NORMAL DISTRIBUTION

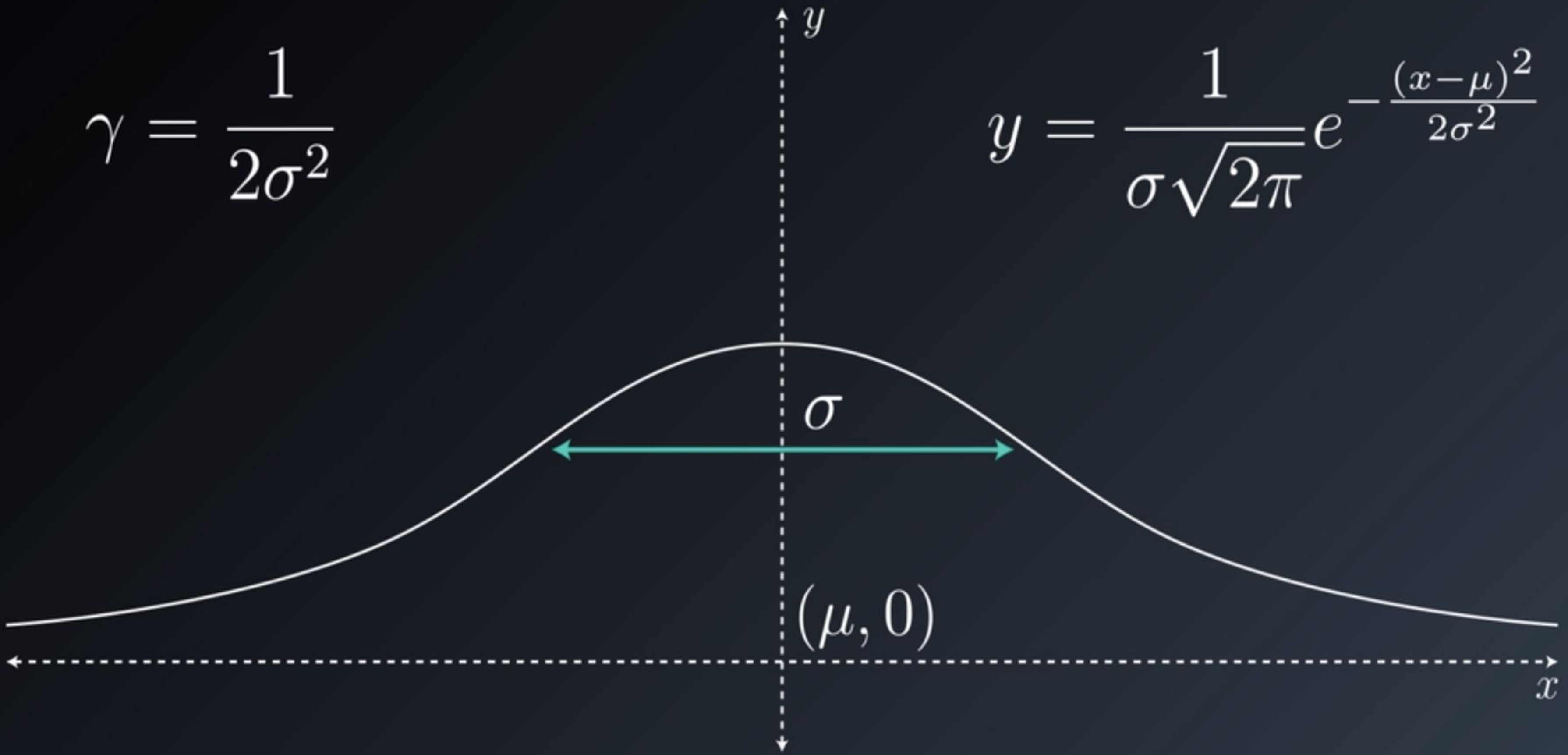
$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)2}{2_{\sigma_2}}}$$



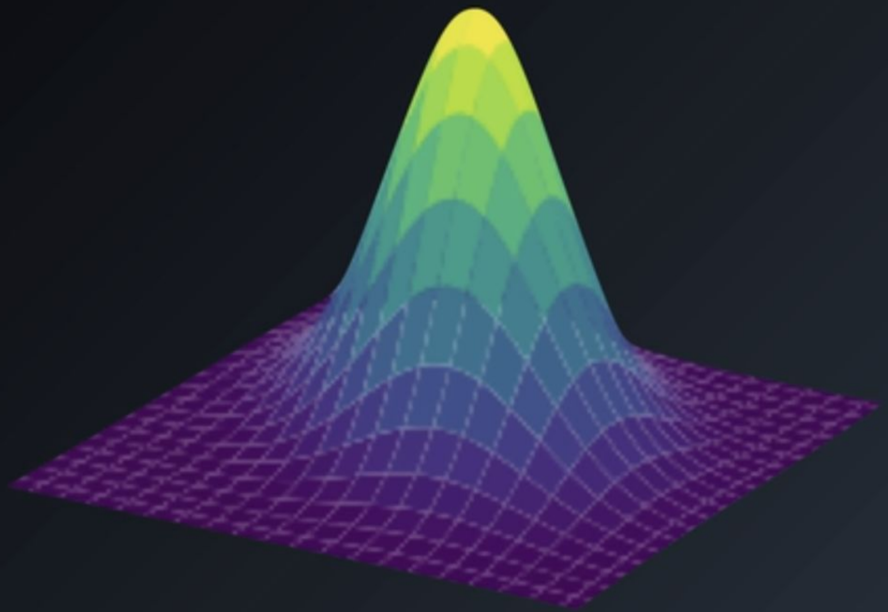
NORMAL DISTRIBUTION

$$\gamma = \frac{1}{2\sigma^2}$$

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



NORMAL DISTRIBUTION



$$z = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)$$