Day-61, Jan-30, 2025 (Magh 17, 2081 BS)

Introduction to Covariance

X : age of a child

Y_1 : height of the child (in)

Age (X)	Height (Y_1)
6	50
7	52
8	55
9	57
10	60
11	62
12	64
13	65
14	67
15	68

Y_2 : grades in a test

Age (X)	Grades (Y_2)
6	5
7	7
8	8
9	3
10	1
11	1
12	6
13	10
14	2
15	7

Y_3 : number of naps per day

Age (X)	Naps per day (Y_3)
6	8
7	7
8	6
9	5
10	4
11	3
12	2
13	1
14	1
15	0

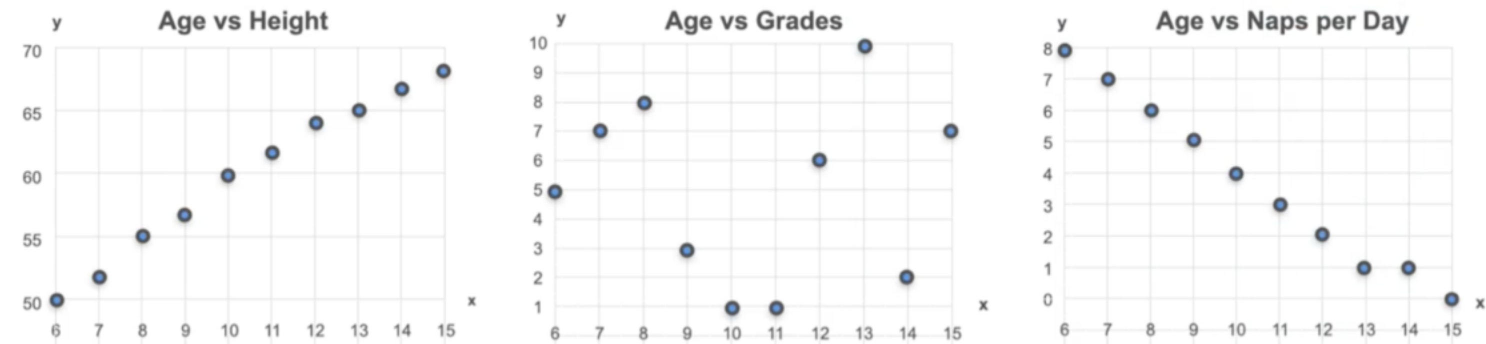
To help us better visualize what's going on

DeepLearning.AI

Introduction to Covariance

X : age of a child

Y_1 : height of the child (in) Y_2 : grades in a test Y_3 : number of naps per day



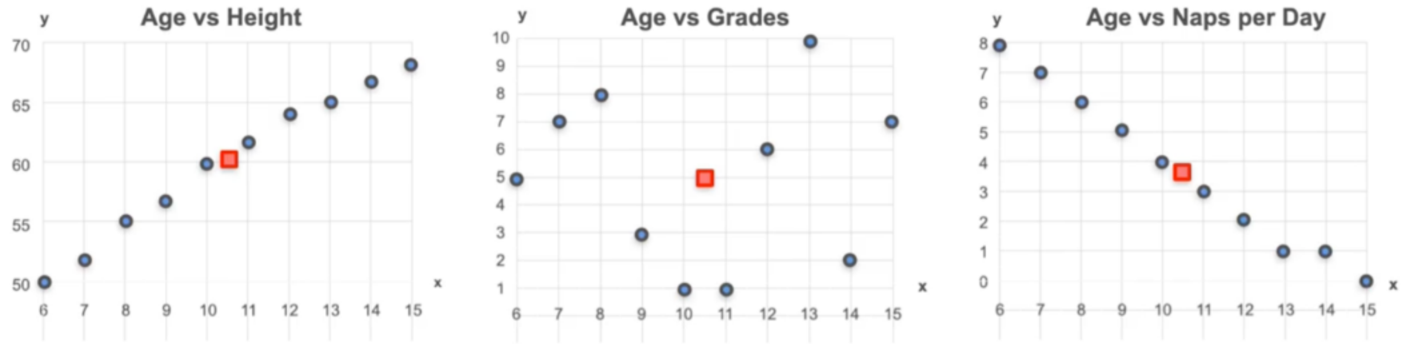
and the third one is like this.

DeepLearning.AI

Co-variance means variance of two variables together

Co-rrelation means having relationship with each other variables.

Mean?



$\mu_x = 10.5$ $\mu_y = 60$

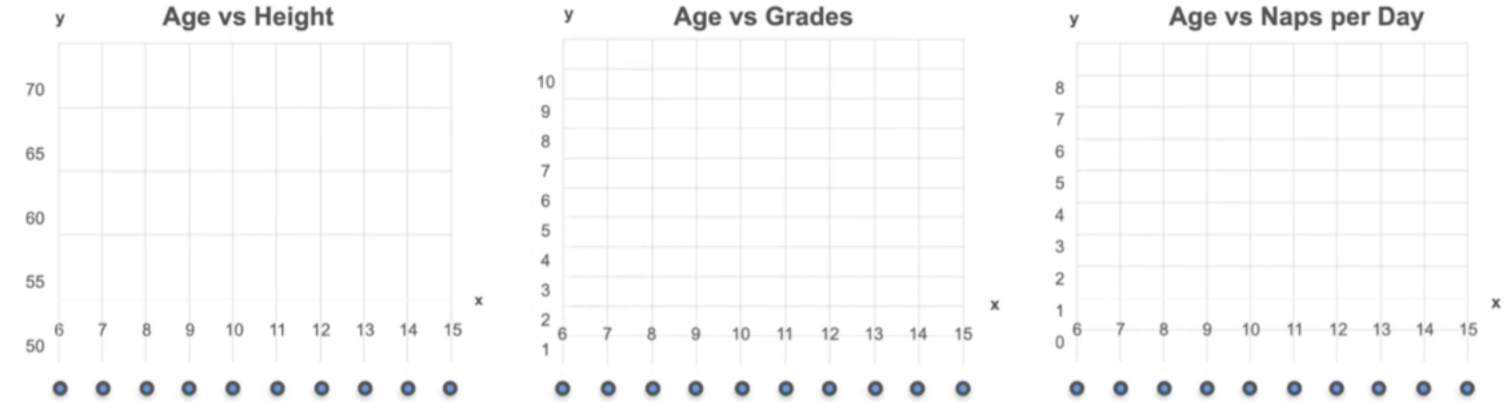
$\mu_x = 10.5$ $\mu_y = 5$

$\mu_x = 10.5$ $\mu_y = 3.7$

Now we can also look at the variances.

DeepLearning.AI

Horizontal (X) Variance



$Var(X) = 9.17$

$Var(X) = 9.17$

$Var(X) = 9.17$

which are 9.17 because it's the same data-set for ages.

DeepLearning.AI

Vertical (Y) Variance

$Var(Y) = 39.56$

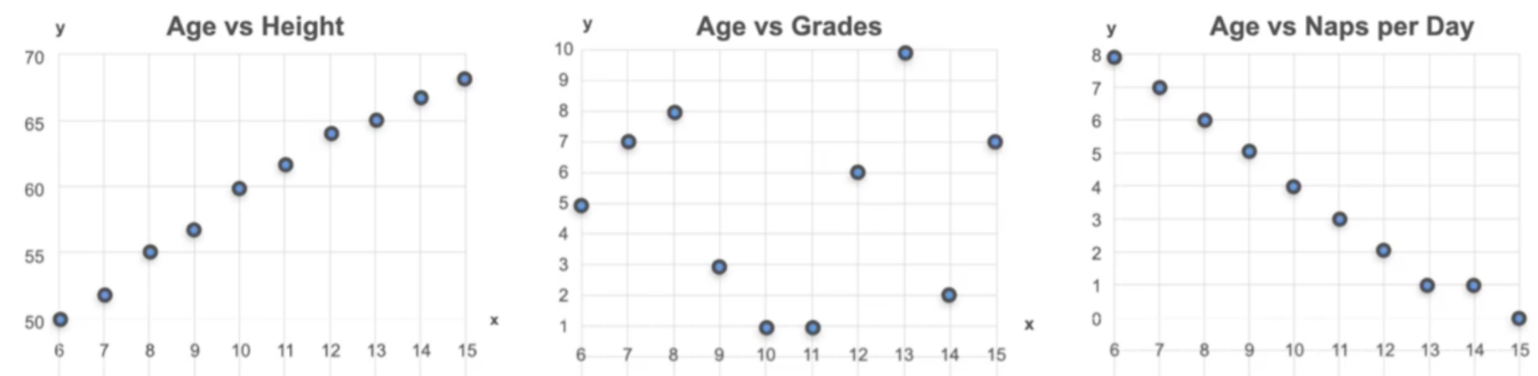
$Var(Y) = 9.78$

$Var(Y) = 7.57$

which are 39.56, 9.78 and 7.57.

DeepLearning.AI

Still no Way To Compare Them



Covariance > 0

Covariance ≈ 0

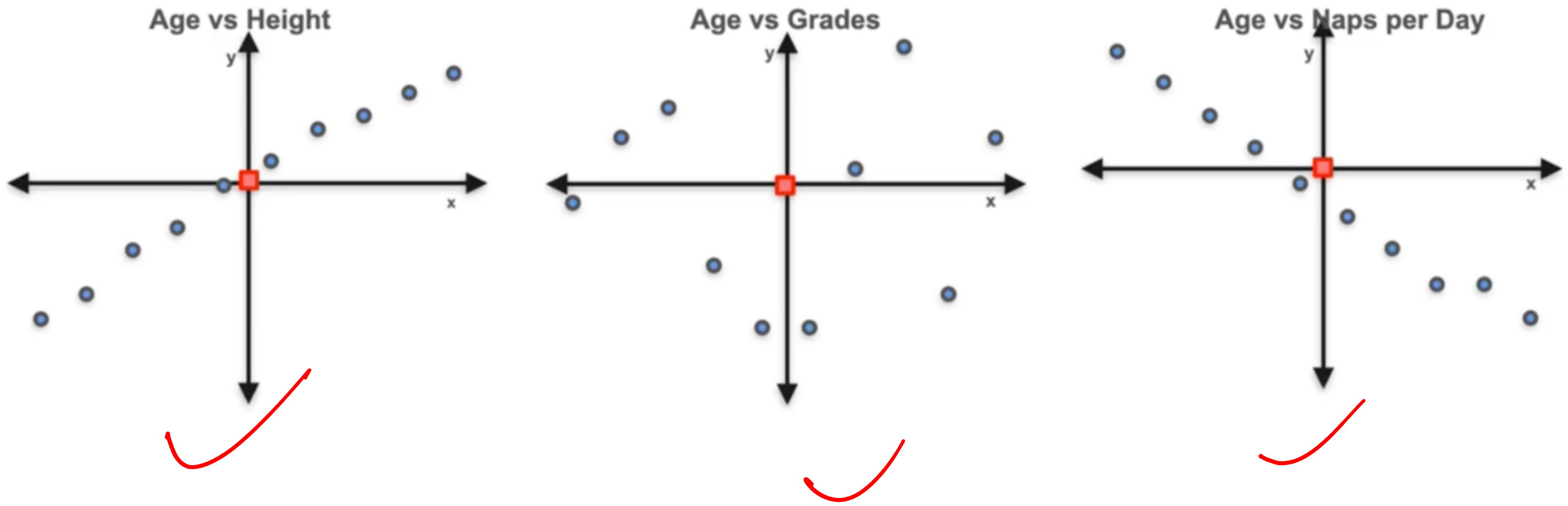
Covariance < 0

the covariance of the second plot is almost zero,

DeepLearning.AI

So, what is the step ?

First Step: Center Them



of X from the x coordinates
and the mean of

DeepLearning.AI

First Step: Center Them

1

$\mu_x = 0$
$\mu_y = 0$
$Var(X) = 1$
$Var(Y) = 1$

2

$\mu_x = 0$
$\mu_y = 0$
$Var(X) = 1$
$Var(Y) = 1$

3

$\mu_x = 0$
$\mu_y = 0$
$Var(X) = 1$
$Var(Y) = 1$

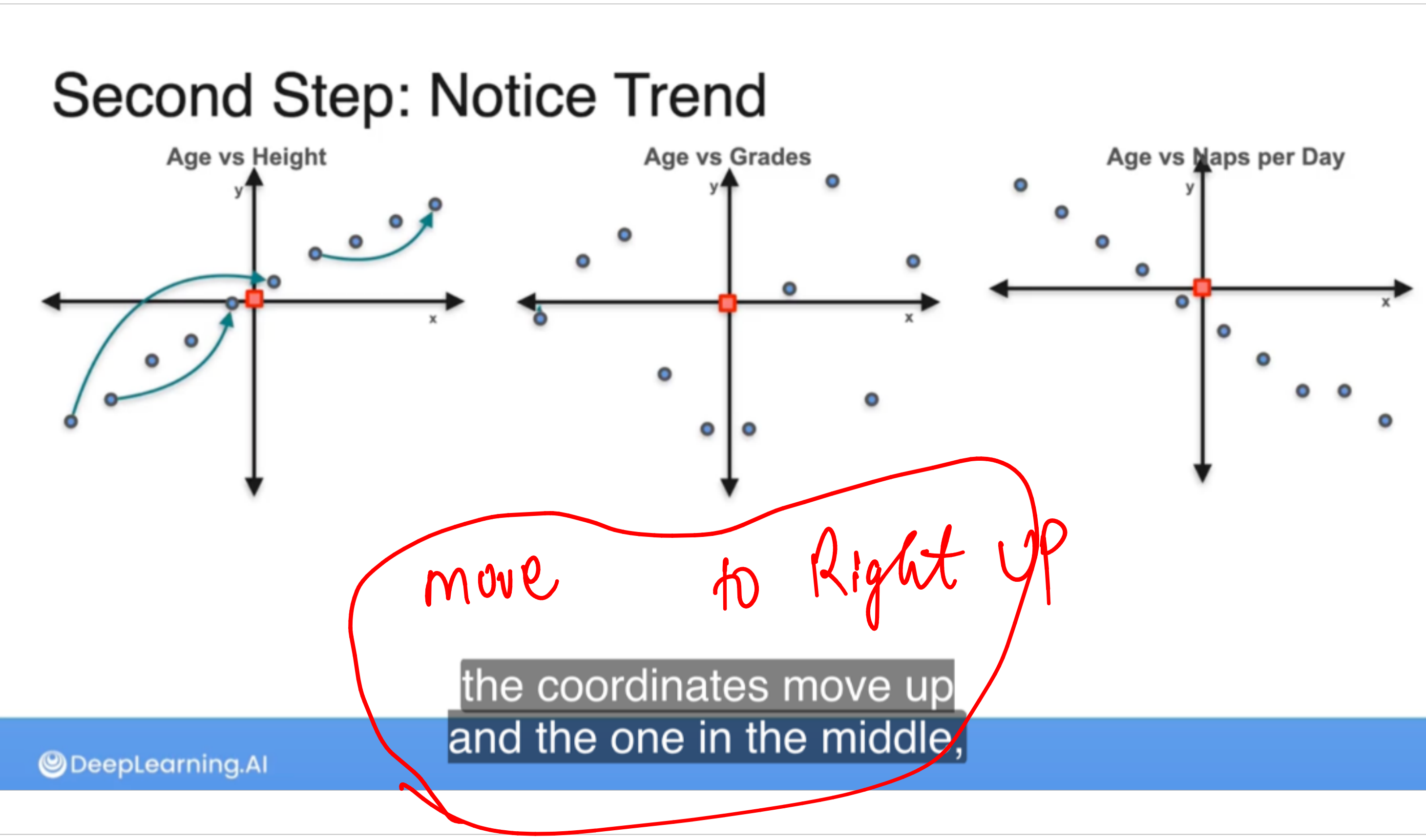
this really nice plots where

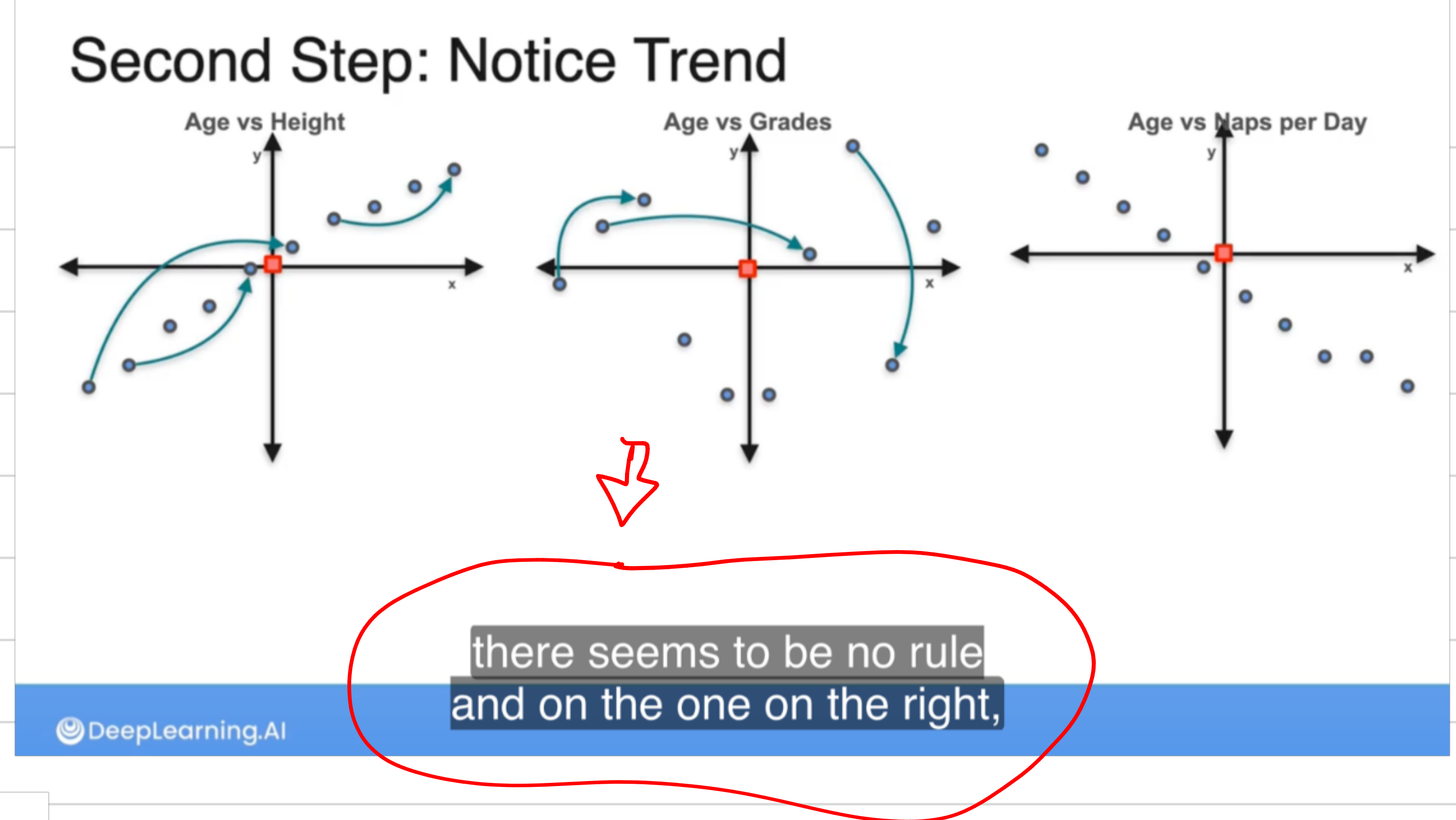
DeepLearning.AI

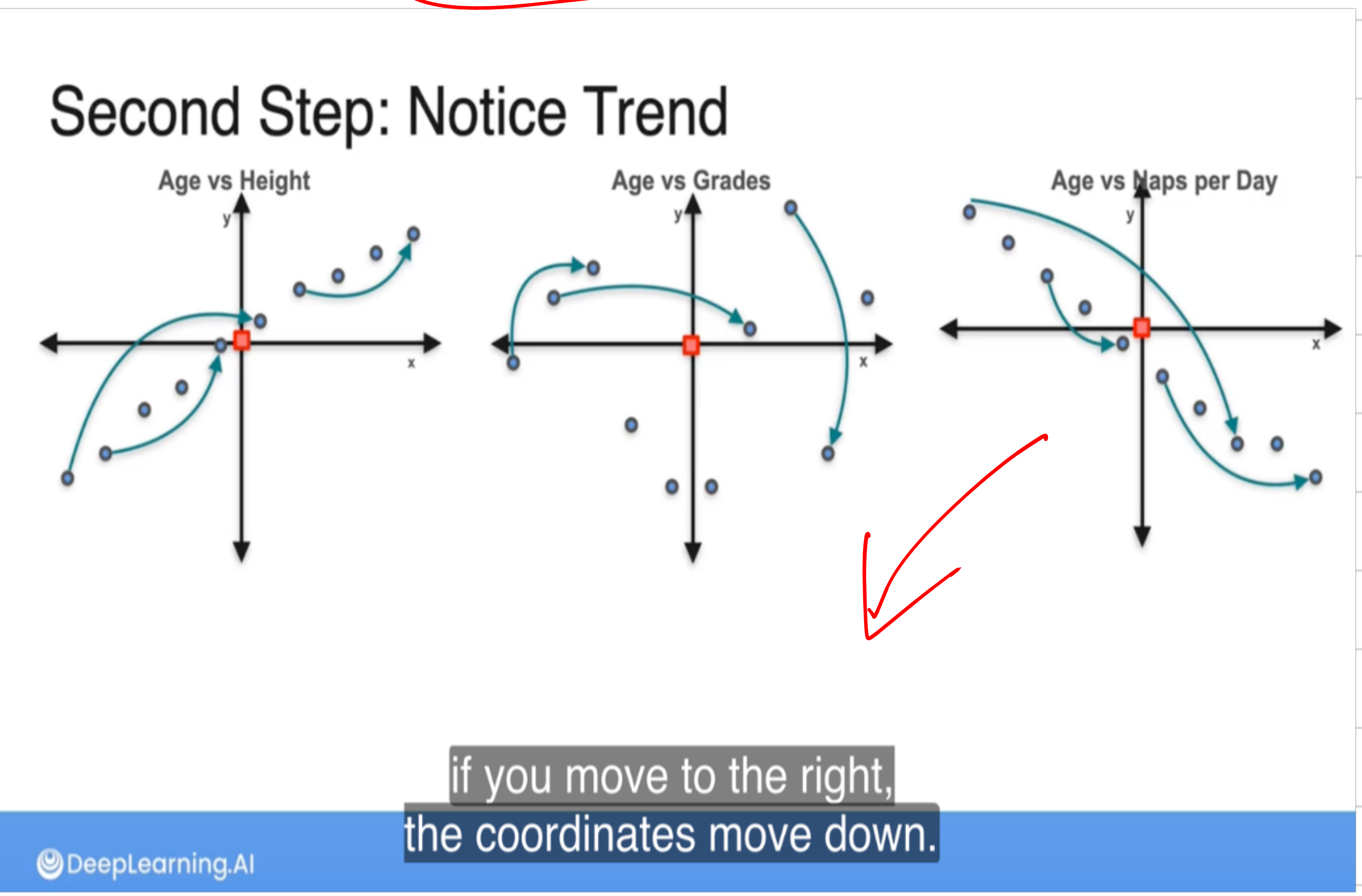
Second Step: Notice Trend

and the lack of
trends in the middle.

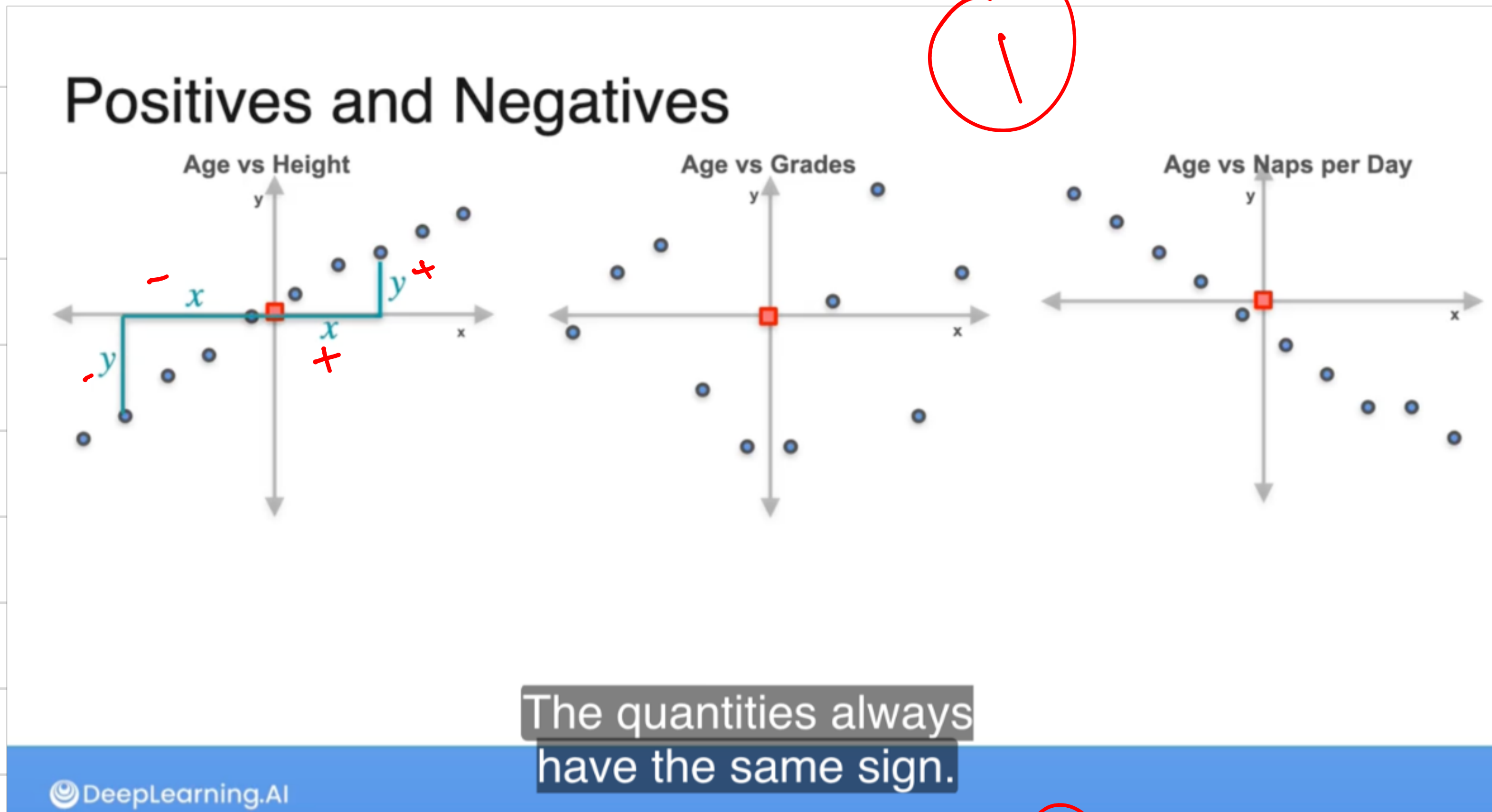
DeepLearning.AI


Second Step: Notice Trend
Age vs Height
Age vs Grades
Age vs Naps per Day
move to Right up
the coordinates move up
and the one in the middle,
DeepLearning.AI

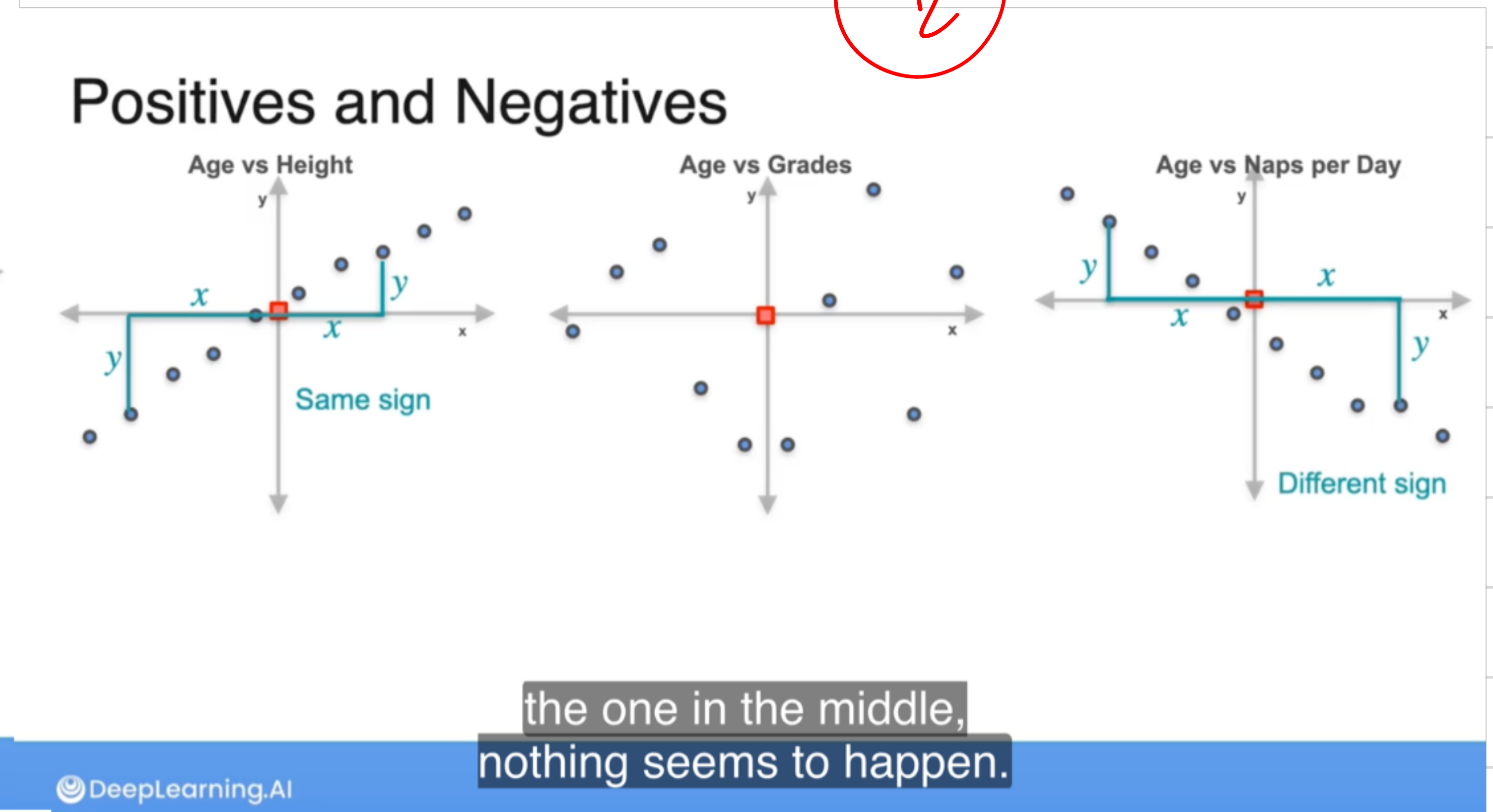

Second Step: Notice Trend
Age vs Height
Age vs Grades
Age vs Naps per Day
there seems to be no rule
and on the one on the right,
DeepLearning.AI


Second Step: Notice Trend
Age vs Height
Age vs Grades
Age vs Naps per Day
if you move to the right,
the coordinates move down.
DeepLearning.AI

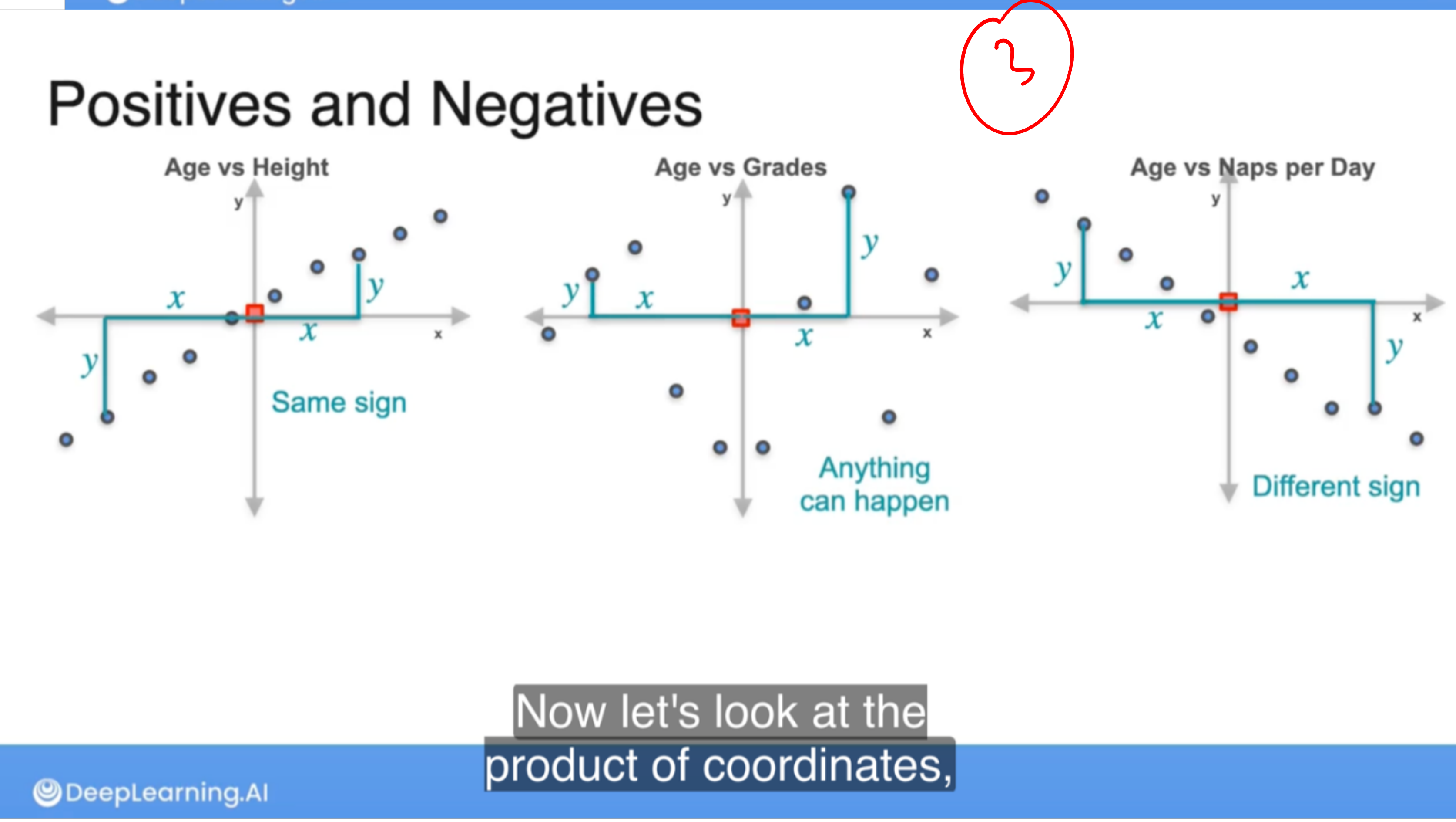
1


Positives and Negatives
Age vs Height
Age vs Grades
Age vs Naps per Day
The quantities always
have the same sign.
DeepLearning.AI

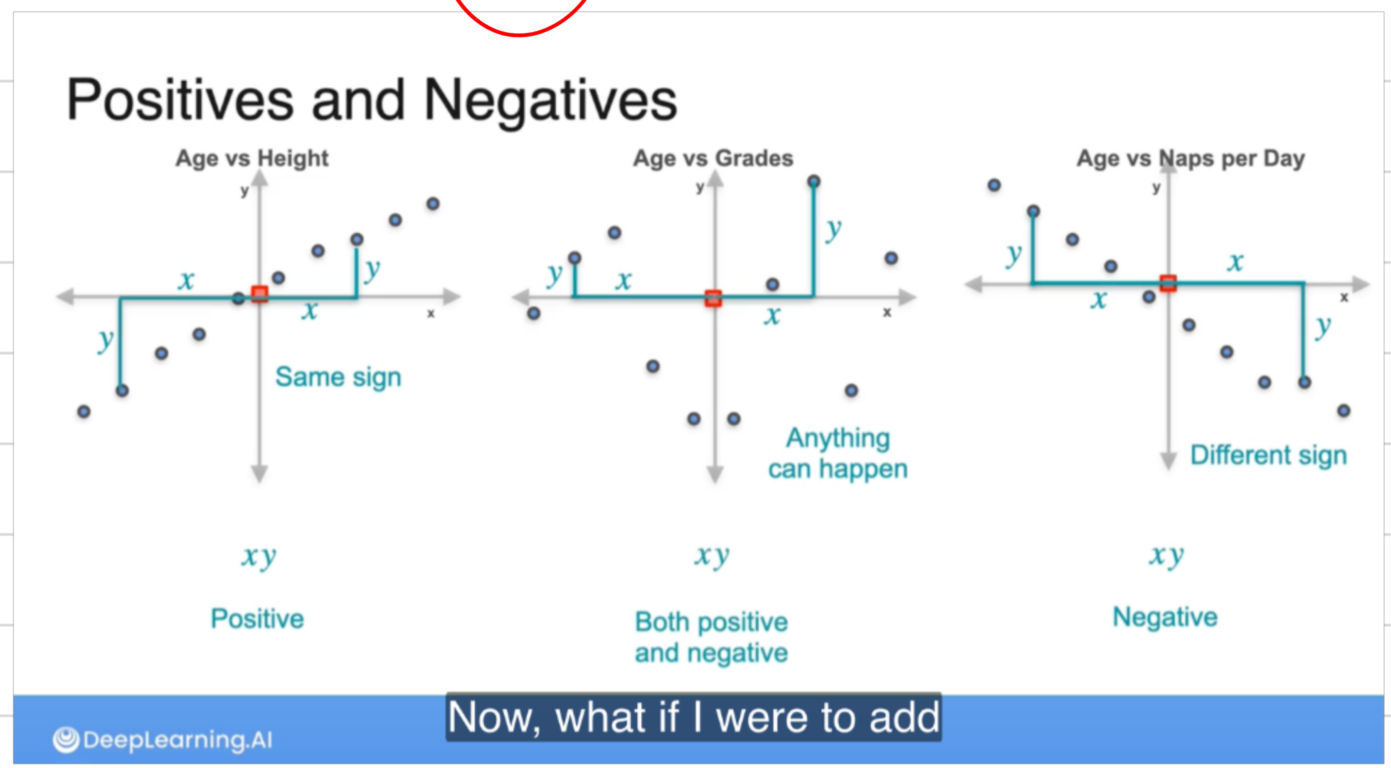
2


Positives and Negatives
Age vs Height
Age vs Grades
Age vs Naps per Day
Same sign
Different sign
the one in the middle,
nothing seems to happen.
DeepLearning.AI

3

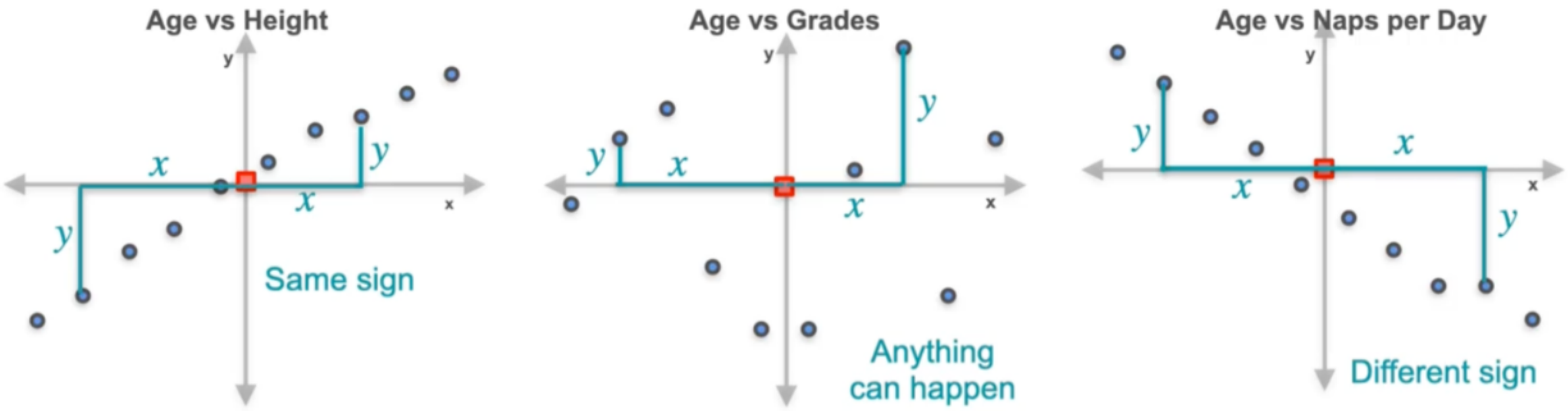

Positives and Negatives
Age vs Height
Age vs Grades
Age vs Naps per Day
Same sign
Anything
can happen
Different sign
Now let's look at the
product of coordinates,
DeepLearning.AI

4


Positives and Negatives
Age vs Height
Age vs Grades
Age vs Naps per Day
Same sign
Anything
can happen
Different sign
xy
Positive
xy
Both positive
and negative
xy
Negative
Now, what if I were to add
DeepLearning.AI

1

Covarian egatives



$\sum xy > 0$ — Positive

$\sum xy \approx 0$ — Both positive and negative

$\sum xy < 0$ — Negative

This sum is called the co-variance.

DeepLearning.AI

2

Covariance

$$Cov(X,Y) = \sum xy$$ Almost…

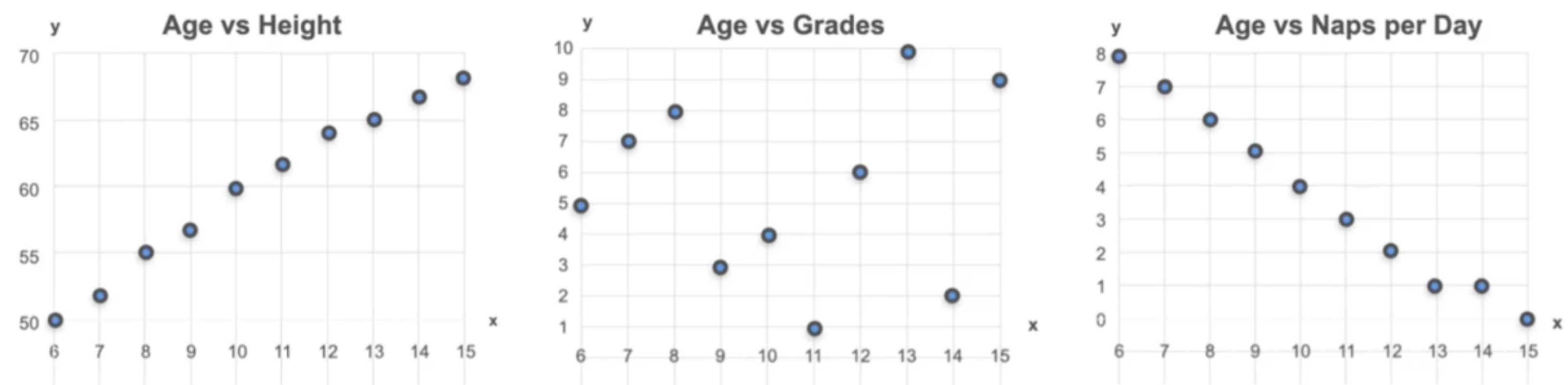
$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

take the average of all these products.

DeepLearning.AI

3

Covariance



$Cov(X,Y) > 0$ $\quad$ $Cov(X,Y) \approx 0$ $\quad$ $Cov(X,Y) < 0$

like age and naps per day.

DeepLearning.AI

4

Covariance Formula

Age vs Height

Covariance > 0

$\mu_x = 10.5 \quad \mu_y = 60$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

$$Cov(X,Y) = \frac{170}{10} = 17 > 0$$

Age (x)	Height (y)	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
6	50	-4.5	-10	45
7	52	-3.5	-8	28
8	55	-2.5	-5	12.5
9	57	-1.5	-3	4.5
10	60	-0.5	0	0
11	62	0.5	2	1
12	64	1.5	4	6
13	65	2.5	5	12.5
14	67	3.5	7	24.5
15	68	4.5	8	36

$\sum = 170$

that when age grows, height grows.

DeepLearning.AI

1

Covariance Formula

Age vs Height

Covariance > 0

$\mu_x = 10.5 \qquad \mu_y = 60$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

$$Cov(X,Y) = \frac{170}{10} = 17 > 0$$

Age (x)	Height (y)	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
6	50	-4.5	-10	45
7	52	-3.5	-8	28
8	55	-2.5	-5	12.5
9	57	-1.5	-3	4.5
10	60	-0.5	0	0
11	62	0.5	2	1
12	64	1.5	4	6
13	65	2.5	5	12.5
14	67	3.5	7	24.5
15	68	4.5	8	36

$\sum = 170$

that when age grows,
height grows.

DeepLearning.AI

2

Covariance Formula

Age vs Naps per Day

Covariance < 0

$\mu_x = 10.5 \qquad \mu_y = 3.7$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

$$Cov(X,Y) = \frac{-74.5}{10} = -7.45 < 0$$

Age (x)	Naps (y)	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
6	8	-4.5	4.3	-19.35
7	7	-3.5	3.3	-11.55
8	6	-2.5	2.3	-5.75
9	5	-1.5	1.3	-1.95
10	4	-0.5	0.3	-0.15
11	3	0.5	-0.7	-0..35
12	2	1.5	-1.7	-2.55
13	1	2.5	-2.7	-6.75
14	1	3.5	-2.7	-9.45
15	0	4.5	-3.7	-16.65

$\sum = -74.5$

That is negative
because as you know,

DeepLearning.AI

3

Covariance Formula

Age vs Grades

Covariance ≈ 0

$\mu_x = 10.5 \qquad \mu_y = 5$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

$$Cov(X,Y) = \frac{1}{10} = 0.1 \quad \approx 0$$

Age (x)	Grades (y)	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
6	5	-4.5	0	0
7	7	-3.5	2	-7
8	8	-2.5	3	-7.5
9	3	-1.5	-2	3
10	1	-0.5	-4	2
11	1	0.5	-4	-2
12	6	1.5	1	1.5
13	10	2.5	5	12.5
14	2	3.5	-3	-10.5
15	7	4.5	2	9

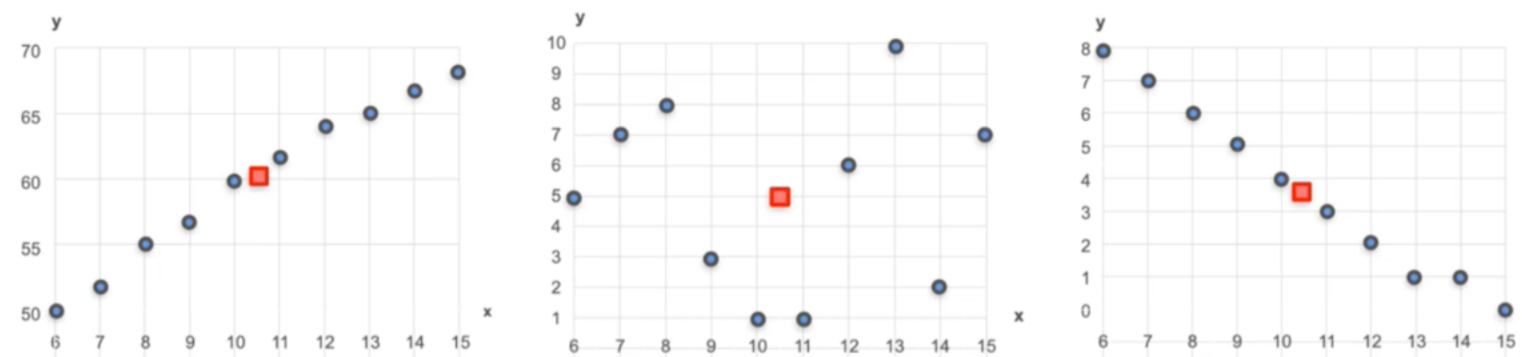
$\sum = 1$

or at least very
little influence for

DeepLearning.AI

4

Comparing Correlations

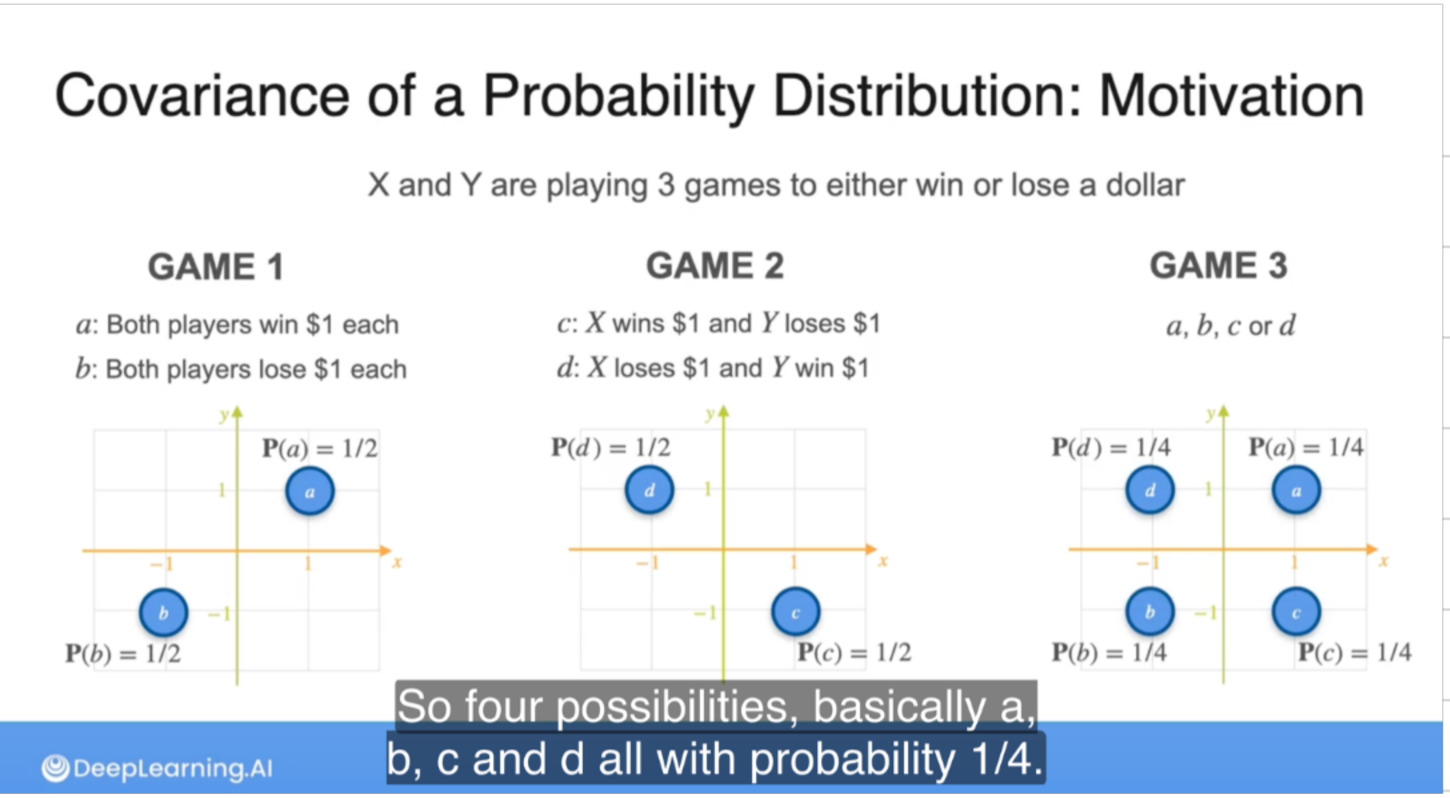


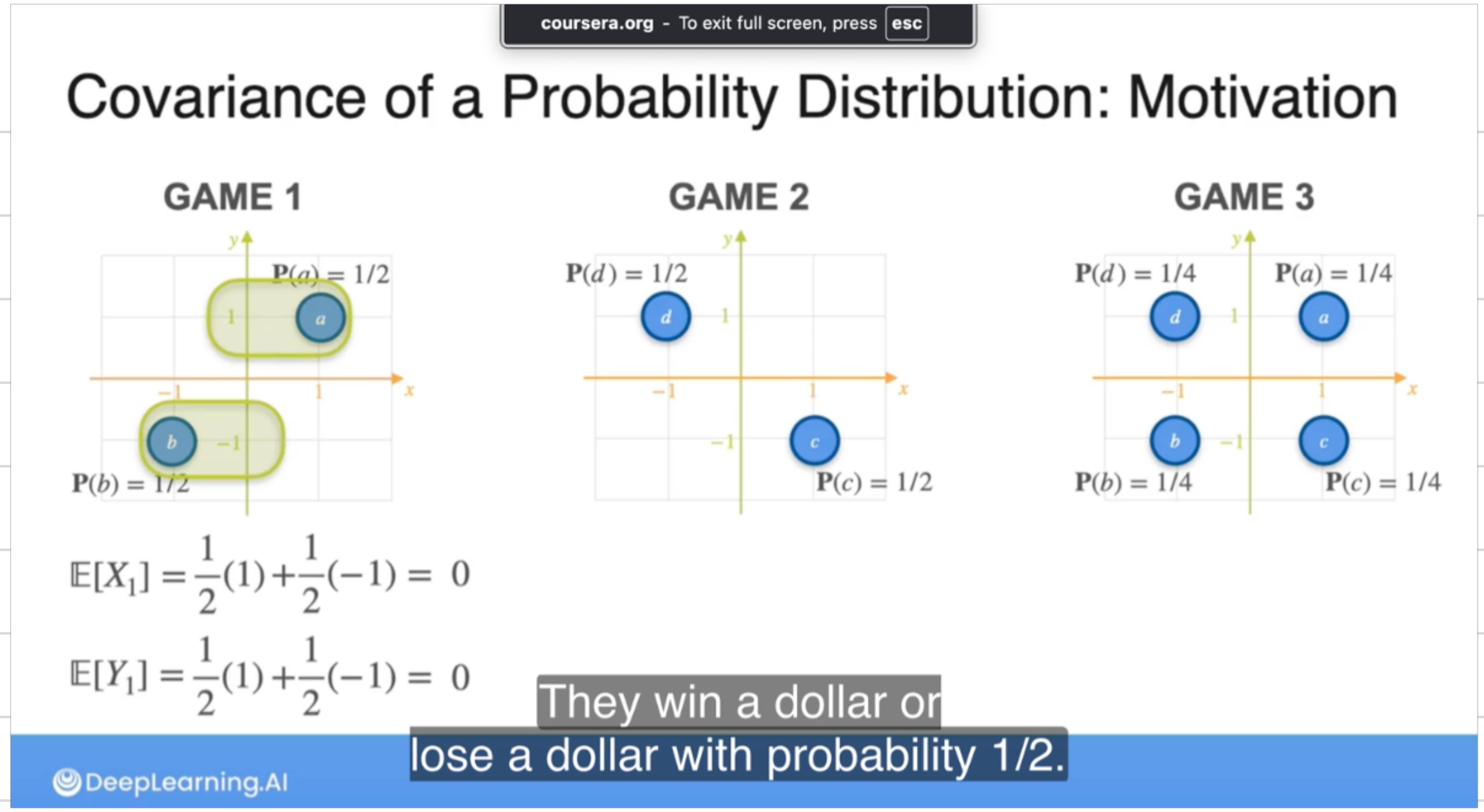
Age vs Height	Age vs Grades	Age vs Naps per Day
Covariance > 0	Covariance ≈ 0	Covariance < 0
$Cov(x,y) = 17$	$Cov(x,y) = 0.1$	$Cov(x,y) = -7.45$

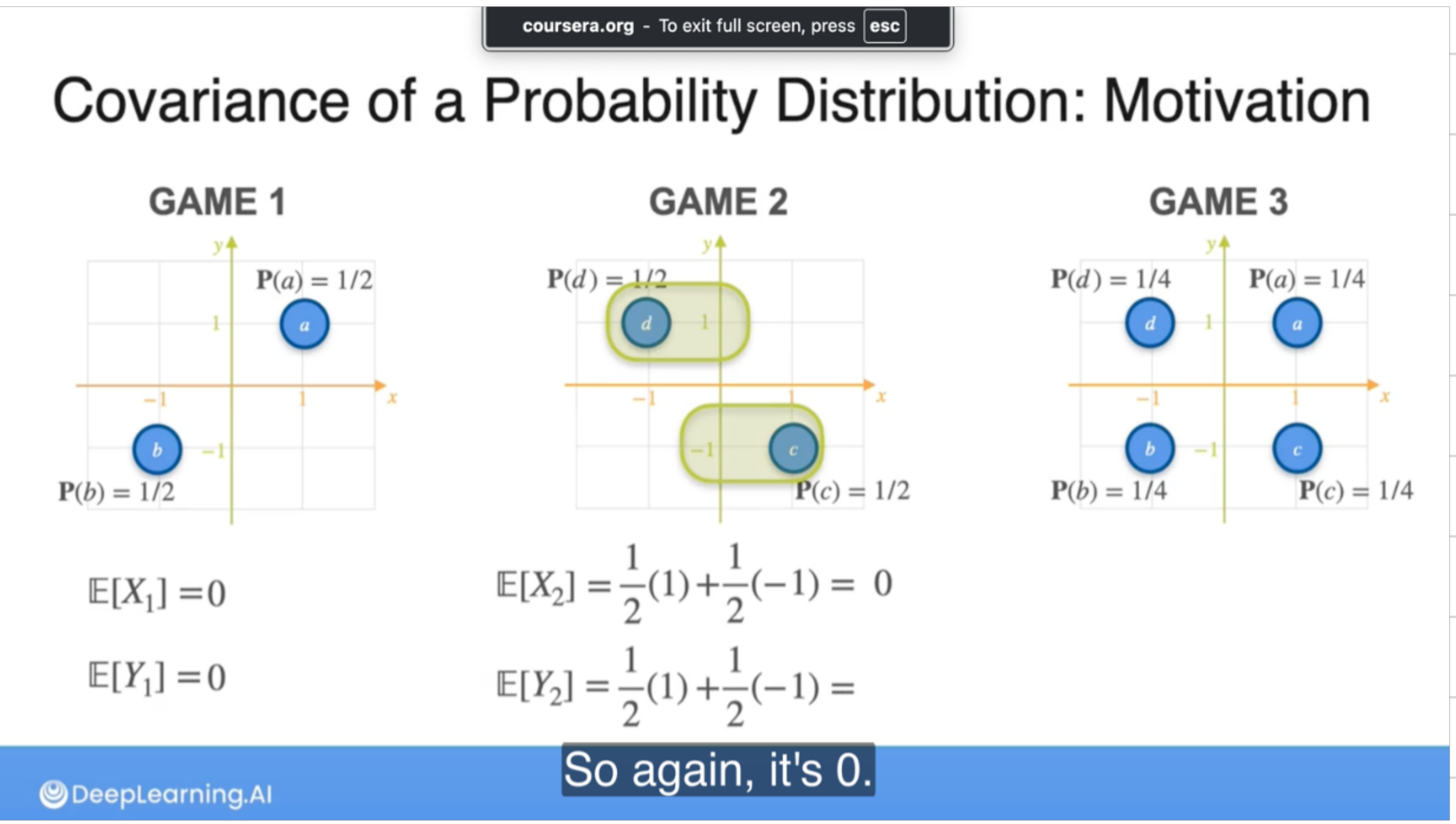
be negatively correlated
have a covariance of -7.45,

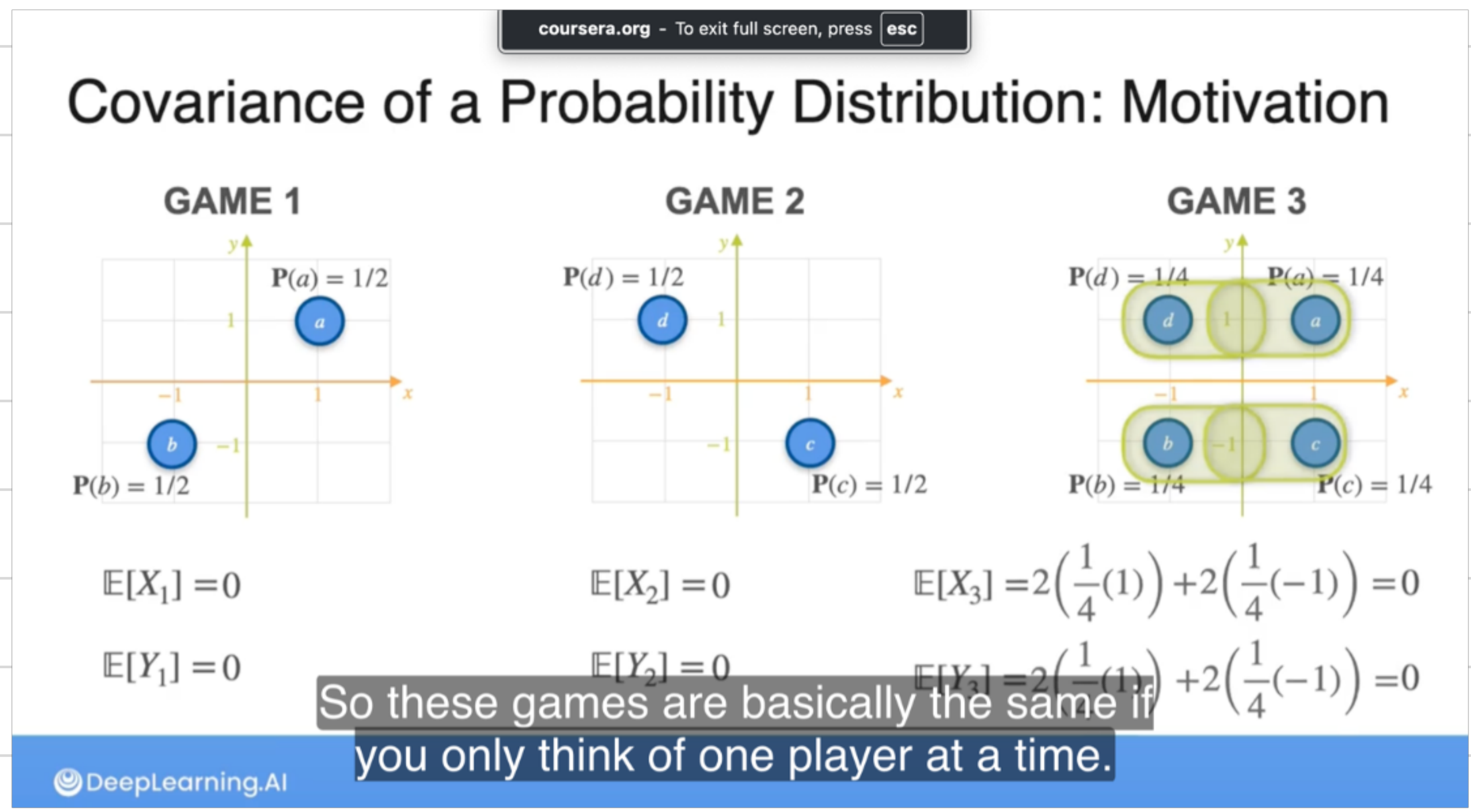
DeepLearning.AI

+ve, -ve and. No Covariance


Covariance of a Probability Distribution: Motivation
X and Y are playing 3 games to either win or lose a dollar
GAME 1
a: Both players win $1 each
b: Both players lose $1 each
GAME 2
c: X wins $1 and Y loses $1
d: X loses $1 and Y win $1
GAME 3
a, b, c or d
P(a) = 1/2
P(b) = 1/2
P(d) = 1/2
P(c) = 1/2
P(d) = 1/4
P(a) = 1/4
P(b) = 1/4
P(c) = 1/4
So four possibilities, basically a, b, c and d all with probability 1/4.
DeepLearning.AI

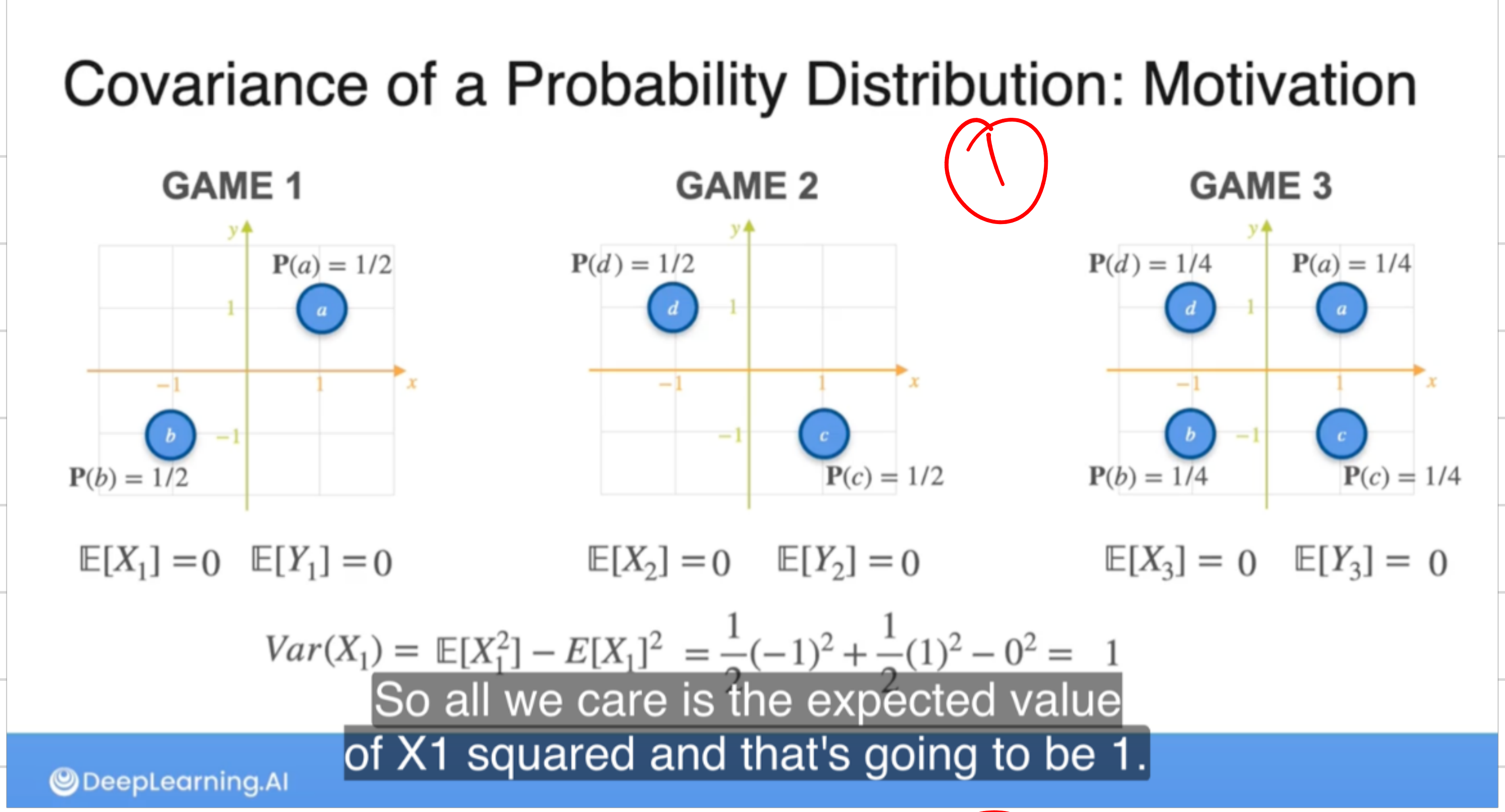

coursera.org - To exit full screen, press esc
Covariance of a Probability Distribution: Motivation
GAME 1
GAME 2
GAME 3
P(a) = 1/2
P(b) = 1/2
P(d) = 1/2
P(c) = 1/2
P(d) = 1/4
P(a) = 1/4
P(b) = 1/4
P(c) = 1/4
$\mathbb{E}[X_1] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$
$\mathbb{E}[Y_1] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$
They win a dollar or lose a dollar with probability 1/2.
DeepLearning.AI


coursera.org - To exit full screen, press esc
Covariance of a Probability Distribution: Motivation
GAME 1
GAME 2
GAME 3
P(a) = 1/2
P(b) = 1/2
P(d) = 1/2
P(c) = 1/2
P(d) = 1/4
P(a) = 1/4
P(b) = 1/4
P(c) = 1/4
$\mathbb{E}[X_1] = 0$
$\mathbb{E}[Y_1] = 0$
$\mathbb{E}[X_2] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$
$\mathbb{E}[Y_2] = \frac{1}{2}(1) + \frac{1}{2}(-1) =$
So again, it's 0.
DeepLearning.AI

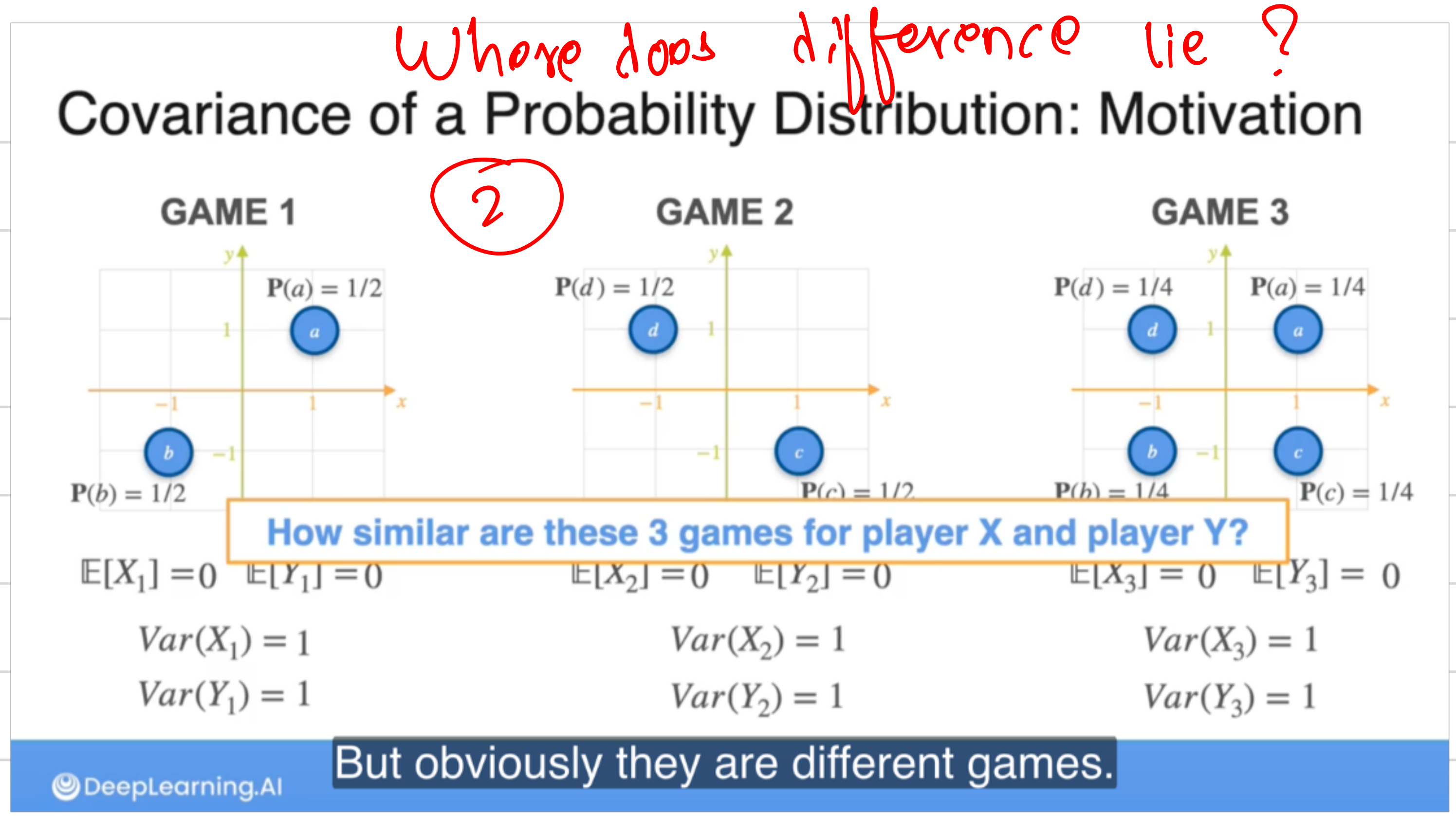

coursera.org - To exit full screen, press esc
Covariance of a Probability Distribution: Motivation
GAME 1
GAME 2
GAME 3
P(a) = 1/2
P(b) = 1/2
P(d) = 1/2
P(c) = 1/2
P(d) = 1/4
P(a) = 1/4
P(b) = 1/4
P(c) = 1/4
$\mathbb{E}[X_1] = 0$
$\mathbb{E}[Y_1] = 0$
$\mathbb{E}[X_2] = 0$
$\mathbb{E}[Y_2] = 0$
$\mathbb{E}[X_3] = 2\left(\frac{1}{4}(1)\right) + 2\left(\frac{1}{4}(-1)\right) = 0$
$\mathbb{E}[Y_3] = 2\left(\frac{1}{4}(1)\right) + 2\left(\frac{1}{4}(-1)\right) = 0$
So these games are basically the same if you only think of one player at a time.
DeepLearning.AI

Where does difference lie ?

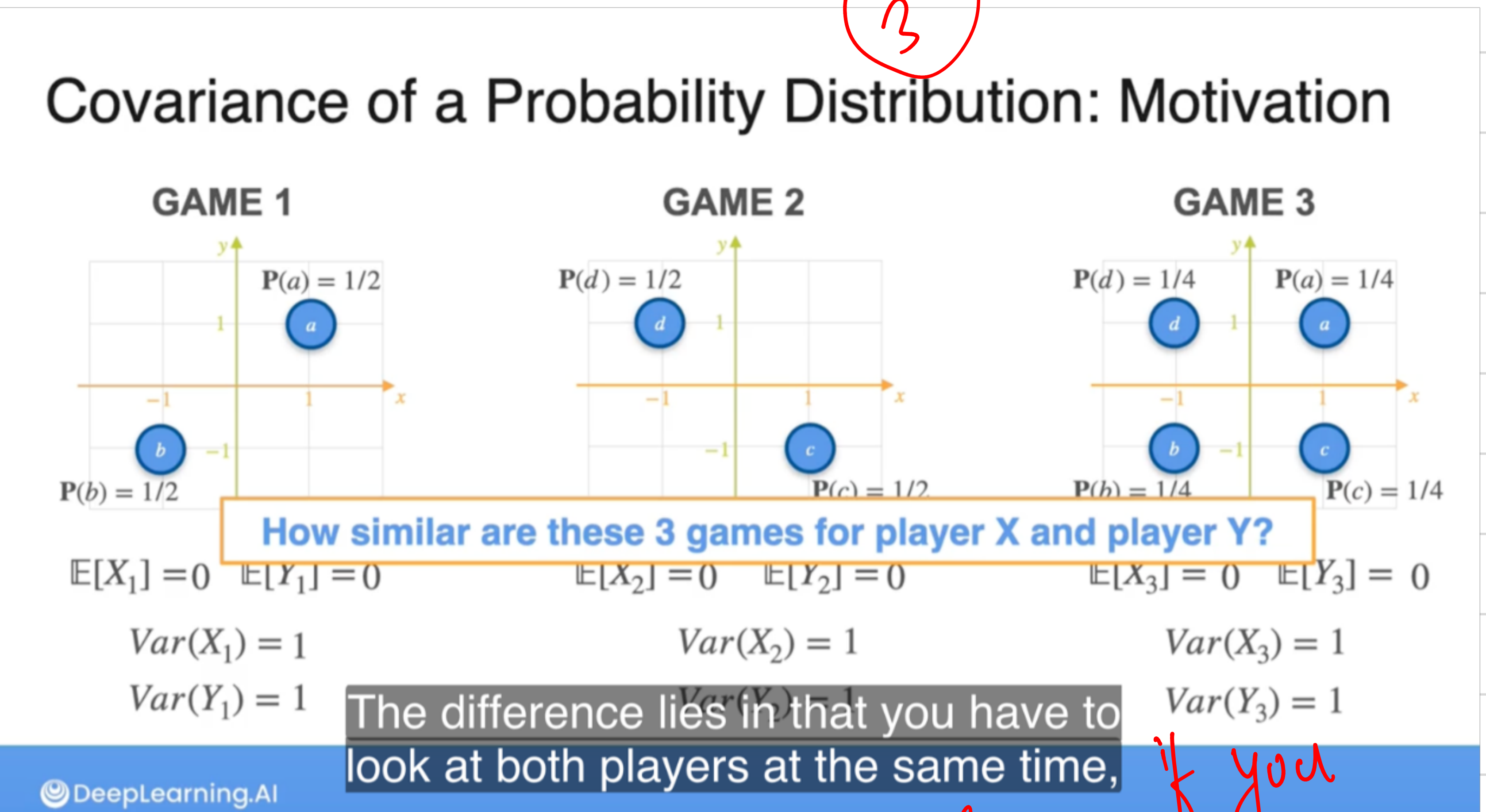
1



2

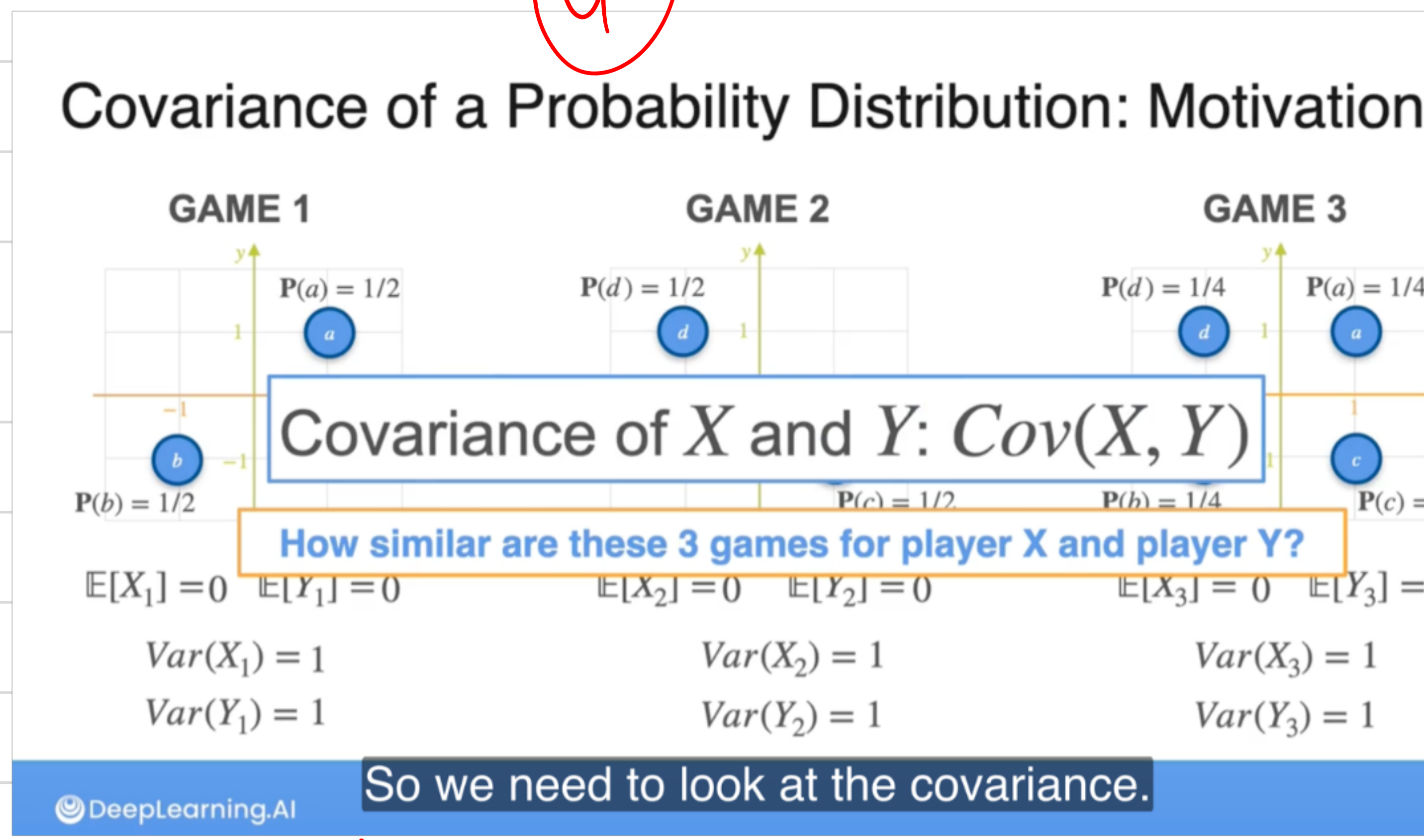


3



if you look one player at a time

4



to know the difference!

Covariance of a Probability Distribution: Motivation

①

Covariance of X and Y: $Cov(X, Y$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

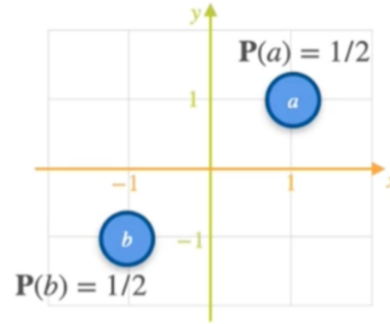
we're going to calculate it for the three games.

DeepLearning.AI

Covariance of a Probability Distribution: Motivation

②

GAME 1



$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

x	y	$(x_i-\mu_x)$	$(y_i-\mu_y)$	$(x_i-\mu_x)(y_i-\mu_y)$
1	1	1	1	1
-1	-1	-1	-1	1

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n} = \frac{2}{2} = 1$$

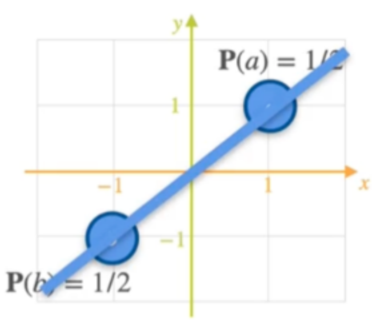
So the covariance of the first game is 1.

DeepLearning.AI

③

Covariance of a Probability Distribution: Motivation

GAME 1



$$\mu_x = \mathbb{E}[X_1] = 0$$

$$\mu_y = \mathbb{E}[Y_1] = 0$$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

x	y	$(x_i-\mu_x)$	$(y_i-\mu_y)$	$(x_i-\mu_x)(y_i-\mu_y)$
1	1	1	1	1
-1	-1	-1	-1	1

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n} = \frac{2}{2} = 1$$

the more player one wins, the more player two wins.

Either both 'H' either both 'S'.

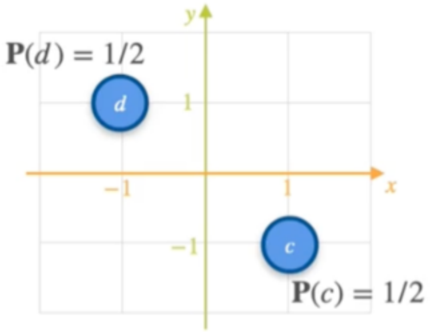
DeepLearning.AI

So, the Cov(X,Y) is +ve that the Game 1 show +ve Covariance.

1

Covariance of a Probability Distribution: Motivation

GAME 2



$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

x	y	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
1	-1	1	-1	-1
-1	1	-1	1	-1

$\mu_x = \mathbb{E}[X_2] = 0$

$\mu_y = \mathbb{E}[Y_2] = 0$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n} = \frac{-2}{2} = -1$$

So this one has a covariance of -1, which is reflected in the fact that

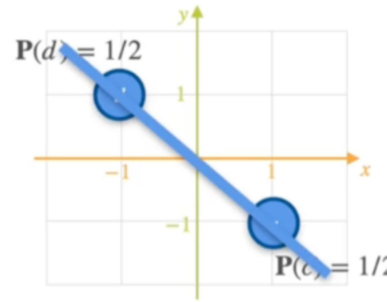
DeepLearning.AI

Game 2 :

2

Covariance of a Probability Distribution: Motivation

GAME 2



$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

x	y	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
1	-1	1	-1	-1
-1	1	-1	1	-1

$\mu_x = \mathbb{E}[X_2] = 0$

$\mu_y = \mathbb{E}[Y_2] = 0$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n} = \frac{-2}{2} = -1$$

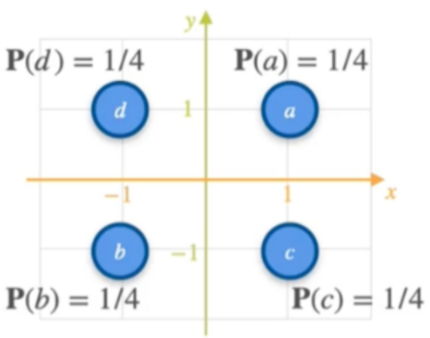
And that means that either one is happy and the other one is sad, or

DeepLearning.AI

3

Covariance of a Probability Distribution: Motivation

GAME 3



$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

x	y	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
1	1	1	1	1
1	-1	-1	1	-1
-1	1	1	-1	-1
-1	-1	-1	-1	1

$\mu_x = \mathbb{E}[X_3] = 0$

$\mu_y = \mathbb{E}[Y_3] = 0$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n} = \frac{0}{4} = 0$$

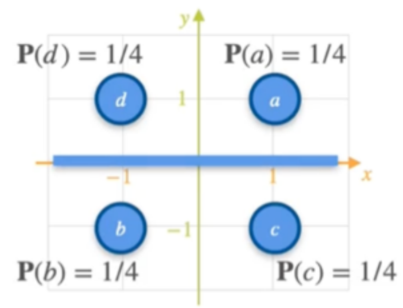
So this covariance is 0.

DeepLearning.AI

4 No pattern All Four Stats

Covariance of a Probability Distribution: Motivation

GAME 3



$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

x	y	$(x_i - \mu_x)$	$(y_i - \mu_y)$	$(x_i - \mu_x)(y_i - \mu_y)$
1	1	1	1	1
1	-1	-1	1	-1
-1	1	1	-1	-1
-1	-1	-1	-1	1

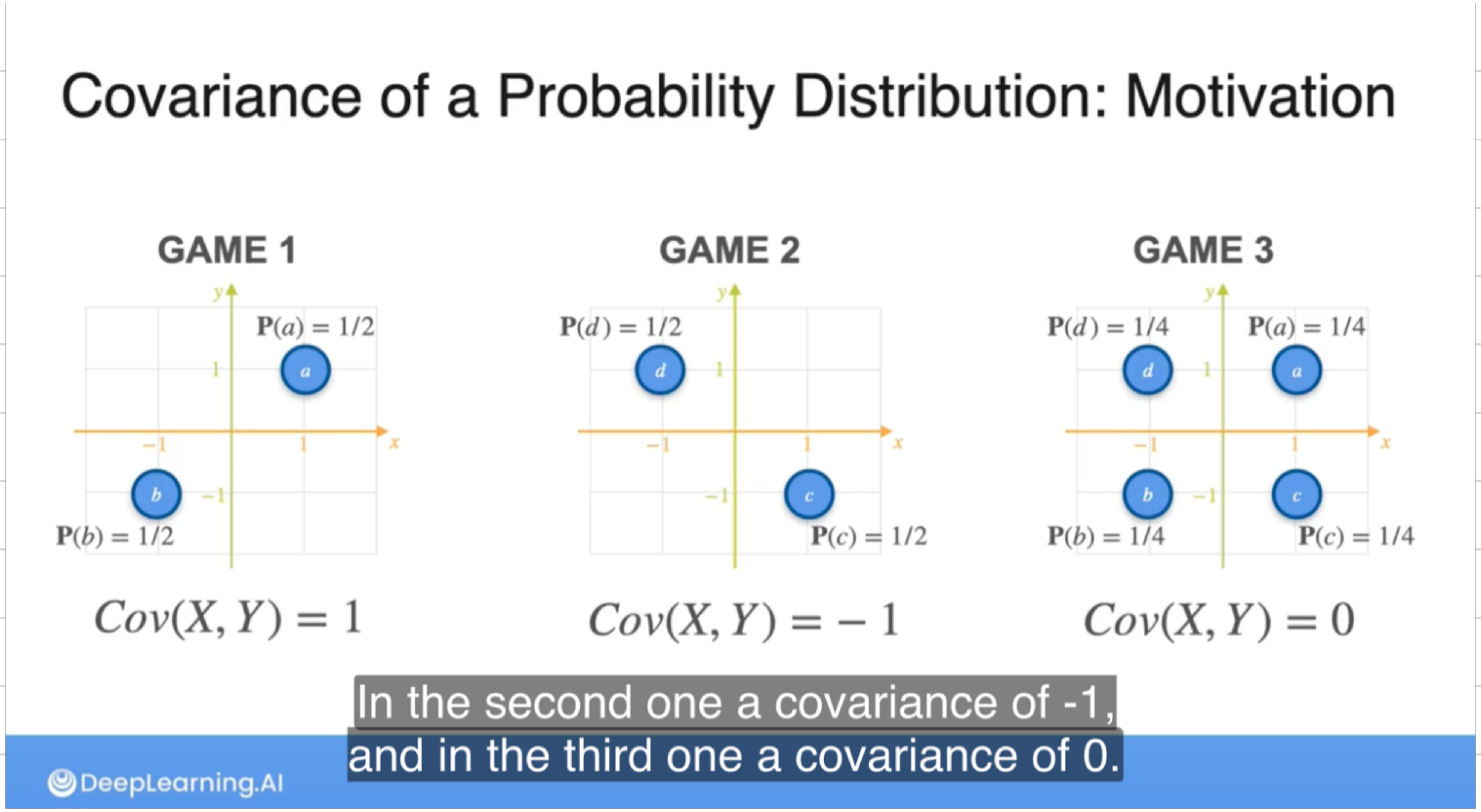
$\mu_x = \mathbb{E}[X_3] = 0$

$\mu_y = \mathbb{E}[Y_3] = 0$

$$Cov(X,Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n} = \frac{0}{4} = 0$$

And that shows the fact that there's not really a pattern among these.

DeepLearning.AI



So, the final Covariance of a Probability Distribution are +ve, 0 and -ve.

So, let's Introduce 3rd Game with three Conditions or states for unequal Probabilities:

Covariance of a Probability Distribution: Motivation

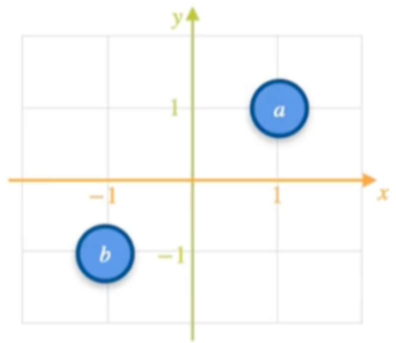
step 1

GAME 4

a: Both players win \$1 each

b: Both players lose \$1 each

c: Neither players wins nor lose anything



Bost players lose a dollar and nothing happens.

DeepLearning.AI

Covariance of a Probability Distribution: Motivation

GAME 4

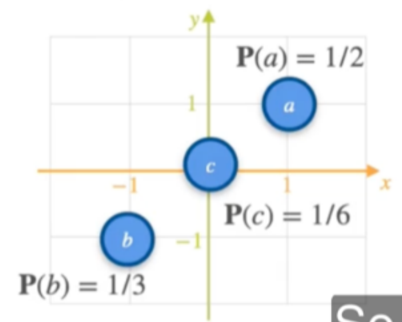
Unequal Probabiliites

a: Both players win \$1 each $\mathbf{P}(a) = 1/2$

b: Both players lose \$1 each $\mathbf{P}(b) = 1/3$

c: Neither players wins nor lose anything $\mathbf{P}(c) = 1/6$

step 2.



So now if we look at only one player, the X player,

DeepLearning.AI

step 3. p ⇒ 1/6 p for each player

Covariance of a Probability Distribution: Motivation

GAME 4

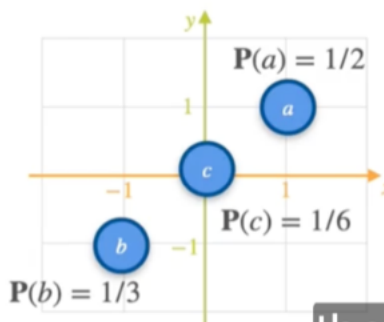
Unequal Probabiliites

a: Both players win \$1 each $\mathbf{P}(a) = 1/2$

b: Both players lose \$1 each $\mathbf{P}(b) = 1/3$

c: Neither players wins nor lose anything $\mathbf{P}(c) = 1/6$

$$Var(X_4) = \sum_{i=1}^{N} (x_i - \mu_x)^2 \cdot \mathbf{P}(x_i)$$



$$\mathbb{E}[X_4] = \frac{1}{6}$$

$$Var(X_4) = \frac{1}{2}\left(1 - \frac{1}{6}\right)^2 + \frac{1}{6}\left(0 - \frac{1}{6}\right)^2 + \frac{1}{3}\left((-1) - \frac{1}{6}\right)^2$$

$$\mathbb{E}[Y_4] = \frac{1}{6}$$

$$= 0.806$$

then multiply with the probabilities and we get 0.806.

DeepLearning.AI

So, $Var(X_4) \Rightarrow 0.806$

$Var(Y_4) = 0.806$

So, what's $Cov(X_4, Y_4) = ?$

use $Cov(X,Y) = p_{XY}(x_i, y_i)(x_i - \mu_x)(y_i - \mu_y)$

Covariance of Probability Distributions

$$Cov(X,Y) = \frac{\sum (x_i-\mu_x)(y_i-\mu_y)}{n} = \boxed{\frac{1}{n}} \sum (x_i-\mu_x)(y_i-\mu_y)$$

equal probabilities

$$Cov(X,Y) = \sum \boxed{p_{XY}(x_i,y_i)} (x_i-\mu_x)(y_i-\mu_y)$$

unequal probabilities

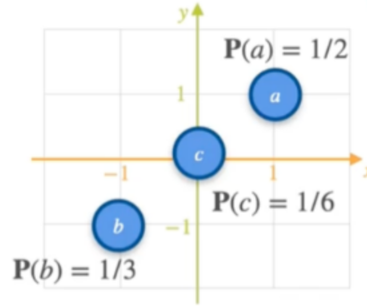
DeepLearning.AI

So this is what happens in general.

Covariance of a Probability Distribution: Motivation

GAME 4

Unequal Probabiliites



$$Cov(X,Y) = \sum p_{XY}(x_i,y_i)\boxed{(x_i-\mu_x)(y_i-\mu_y)} = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$= \frac{1}{2}\boxed{\left(1-\frac{1}{6}\right)^2} + \frac{1}{6}\left(0-\frac{1}{6}\right)^2 + \frac{1}{3}(-1) - \frac{1}{6}\right)^2$$

$Cov(X,Y) = 0.806$

$\mu_x = \mu_y = \mathbb{E}[X_4] = \mathbb{E}[Y_4] = 1/6$

$Var(X_4) = Var(Y_4) = 0.806$

DeepLearning.AI

the covariance XY is 0.806.

Ex-Xx.

$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$

$Cov(X,Y) \Rightarrow 18.014 - 4.903 \times 5.280$

$\Rightarrow -7.878$

-ve Covariance → Slope Downward.

Next is Co-variance Matrix ! .

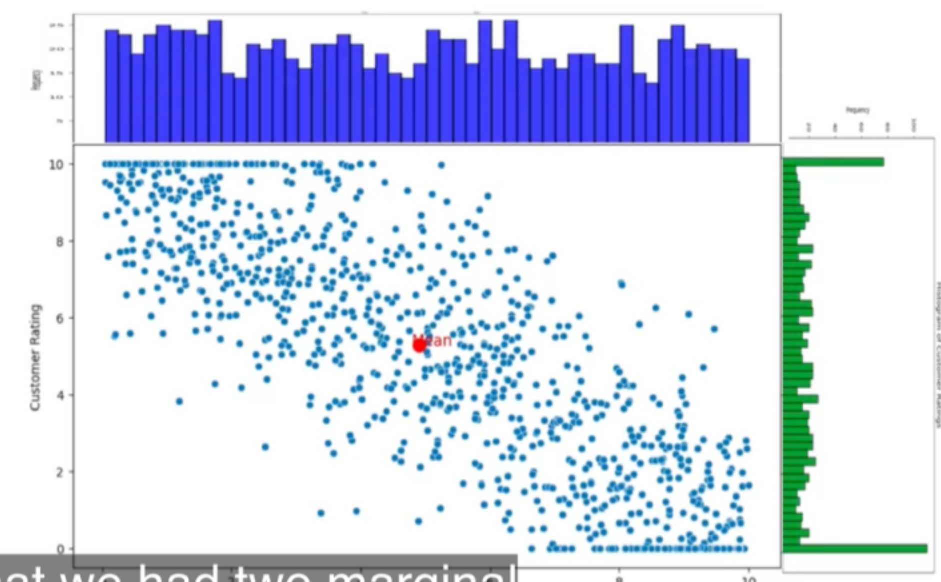
Covariance?

Ex-XX

$\mathbb{E}(X) = 4.903$

$\mathbb{E}(Y) = 5.280$

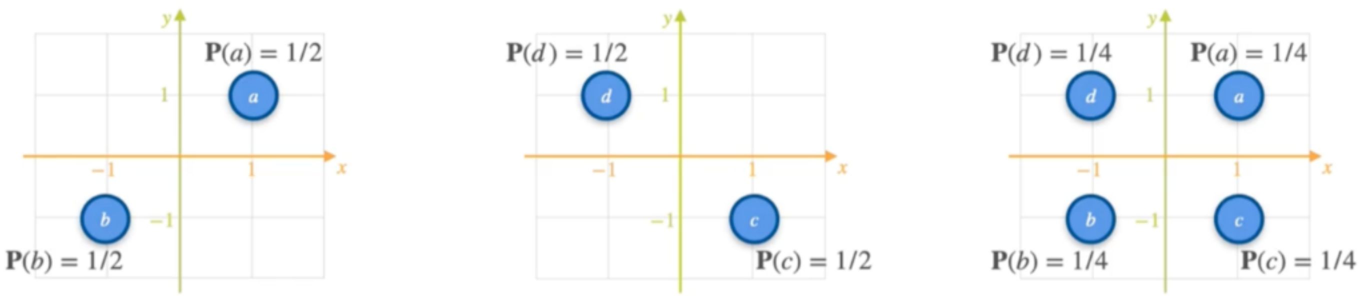
$\mathbb{E}(XY) = 18.014$



Recall that we had two marginal distributions, X and Y and

DeepLearning.AI

Covariance Matrix



$Var(X) = Var(Y) = 1$ $\quad Cov(X,Y) = 1$

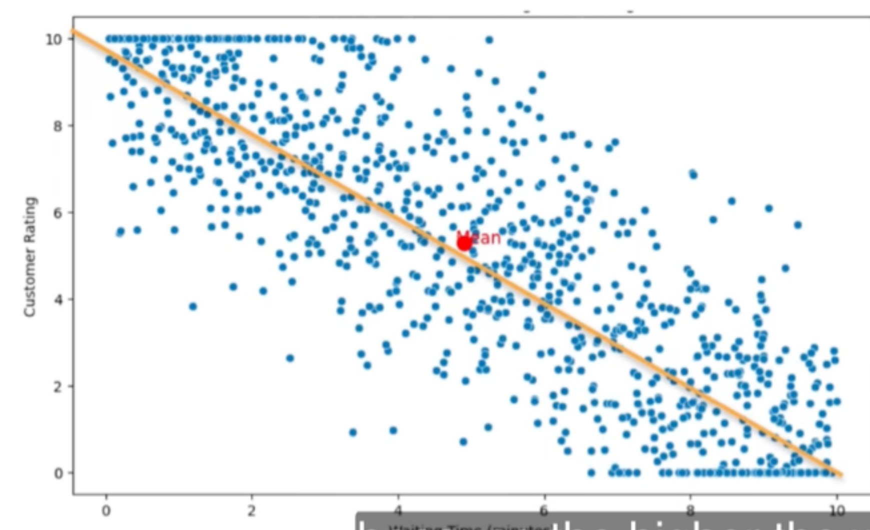
$Var(X) = Var(Y) = 1$ $\quad Cov(X,Y) = -1$

$Var(X) = Var(Y) = 1$ $\quad Cov(X,Y) = 0$

the same expectations and
covariance of 1 minus 1 and zero.

DeepLearning.AI

Covariance Matrix



$Var(X) = 8.526$

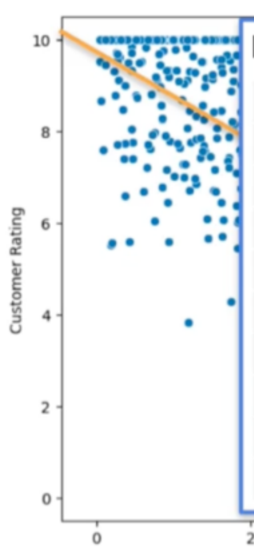
$Var(Y) = 10.163$

$Cov(X,Y) = -7.878$

because the higher the waiting time,
the lower the ratings.

DeepLearning.AI

Covariance Matrix



Dataset with 5 Variables:

V	W	X	Y	Z

$Var(V)$ $\quad Cov(V,W)$

$Var(W)$ $\quad Cov(V,X)$

$Var(X)$ $\quad Cov(V,Y)$

$Var(Y)$ $\quad Cov(V,Z)$

$Var(Z)$ $\quad Cov(W,X)$

$\vdots$

$Cov(Y,Z)$

$ar(X) = 8.526$

$ar(Y) = 10.163$

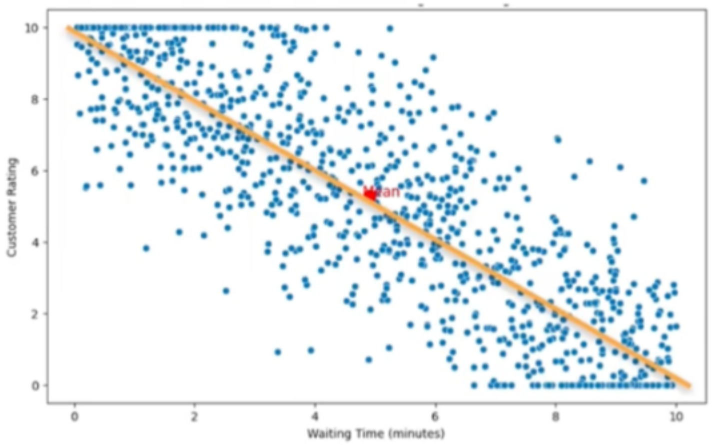
$ov(X,Y) = -7.878$

So we would have ten pairs over here and
that's going to be the covariance matrix.

DeepLearning.AI

Covariance Matrix

1



$Var(X) = 8.526$

$Var(Y) = 10.163$

$Cov(X, Y) = -7.878$

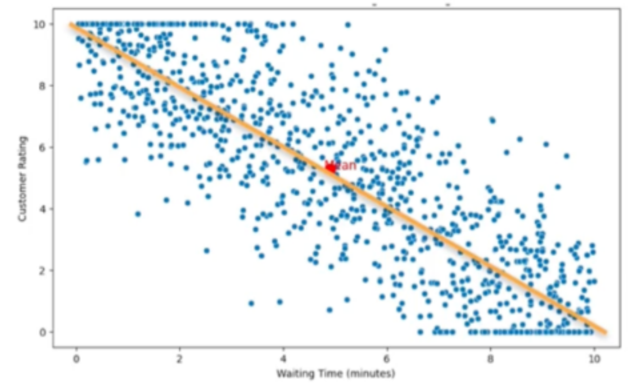
	X	Y
X	$Var(X)$	$Cov(X, Y)$
Y	$Cov(X, Y)$	$Var(Y)$

the covariances on the antidiagonal.

DeepLearning.AI

2

Covariance Matrix



$Var(X) = 8.526$

$Var(Y) = 10.163$

$Cov(X, Y) = -7.878$

	X	Y
X	$Var(X)$	$Cov(X, Y)$
Y	$Cov(X, Y)$	$Var(Y)$

8.526	−7.878
−7.878	10.163

$$\begin{bmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{bmatrix}$$

$$\begin{bmatrix} 8.534 & -7.878 \\ -7.878 & 10.173 \end{bmatrix}$$

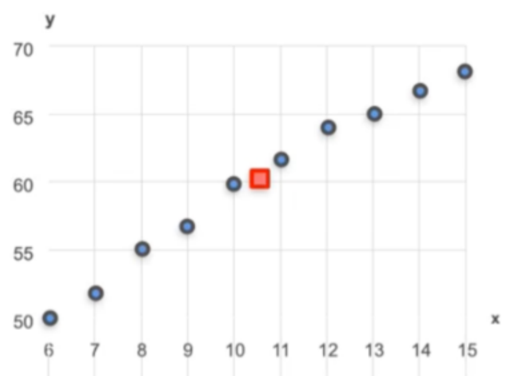
Covariance Matrix

Covariance matrix is very, very useful in statistics and also in machine learning.

DeepLearning.AI

coursera.org - To exit full screen, press esc

Covariance Matrix



Age vs Height

$Var(X) = 9.17$

$Var(Y) = 39.56$

$Cov(X, Y) = 17$

$$\begin{bmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{bmatrix}$$

$$\begin{bmatrix} 9.17 & 17 \\ 17 & 39.56 \end{bmatrix}$$

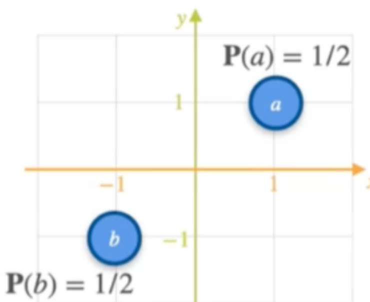
DeepLearning.AI

Matrix with Diagonal Var and antidiagonal Cov are Covariance Matrix

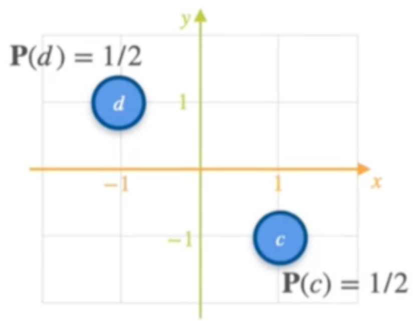
$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$

(1)

Covariance Matrix



$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



$Var(X) = 1$

$Var(Y) = 1$

$Cov(X,Y) = 1$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$Var(X) = 1$

$Var(Y) = 1$

$Cov(X,Y) = -1$

They are these two for games 1 and 2.

DeepLearning.AI

(2)

Covariance of a Joint Continuous Distribution

Dataset with 3 Variables:

X	Y	Z

$Var(X)$ $\quad Cov(X,Y)$

$Var(Y)$ $\quad Cov(X,Z)$

$Var(Z)$ $\quad Cov(Y,Z)$

	X	Y	Z
X	$Var(X)$	$Cov(X,Y)$	$Cov(X,Z)$
Y	$Cov(X,Y)$	$Var(Y)$	$Cov(Y,Z)$
Z	$Cov(X,Z)$	$Cov(Y,Z)$	$Var(Z)$

In the coordinates X set we have the covariance of X set, etc.

DeepLearning.AI

(3)

Covariance of a Joint Continuous Distribution

Dataset with 5 Variables:

V	W	X	Y	Z

$Var(V)$ $\quad Cov(V,W)$

$Var(W)$ $\quad Cov(V,X)$

$Var(X)$ $\quad Cov(V,Y)$

$Var(Y)$ $\quad Cov(V,Z)$

$Var(Z)$ $\quad Cov(W,X)$

$\vdots$

$Cov(Y,Z)$

	V	W	X	Y	Z
V	$Var(V)$	$Cov(V,W)$	$Cov(V,X)$	$Cov(V,Y)$	$Cov(V,Z)$
W	$Cov(V,W)$	$Var(W)$	$Cov(W,X)$	$Cov(W,Y)$	$Cov(W,Z)$
X	$Cov(V,X)$	$Cov(W,X)$	$Var(X)$	$Cov(X,Y)$	$Cov(X,Z)$
Y	$Cov(V,Y)$	$Cov(W,Y)$	$Cov(X,Y)$	$Var(Y)$	$Cov(Y,Z)$
Z	$Cov(V,Z)$	$Cov(W,Z)$	$Cov(X,Z)$	$Cov(Y,Z)$	$Var(Z)$

covariance of the row variable and the column variable.

DeepLearning.AI

(4)

Covariance of a Joint Continuous Distribution

$\Sigma =$

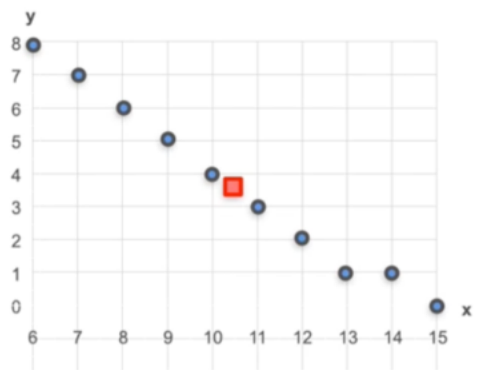
	V	W	X	Y	Z
V	$Var(V)$	$Cov(V,W)$	$Cov(V,X)$	$Cov(V,Y)$	$Cov(V,Z)$
W	$Cov(V,W)$	$Var(W)$	$Cov(W,X)$	$Cov(W,Y)$	$Cov(W,Z)$
X	$Cov(V,X)$	$Cov(W,X)$	$Var(X)$	$Cov(X,Y)$	$Cov(X,Z)$
Y	$Cov(V,Y)$	$Cov(W,Y)$	$Cov(X,Y)$	$Var(Y)$	$Cov(Y,Z)$
Z	$Cov(V,Z)$	$Cov(W,Z)$	$Cov(X,Z)$	$Cov(Y,Z)$	$Var(Z)$

Covariance Matrix

This matrix is going to be called sigma, the covariance matrix.

DeepLearning.AI

Correlation Coefficient

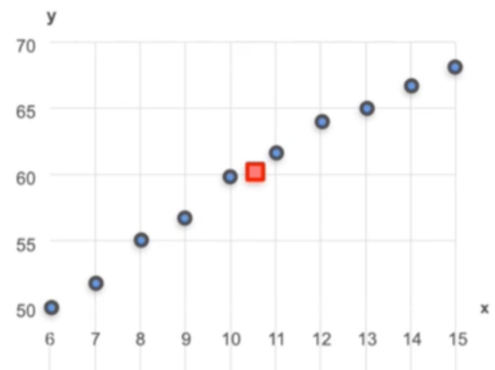


Age vs Naps per Day

$Var(X) = 9.17$

$Var(Y) = 7.57$

$\boxed{Cov(X, Y) = -7.45}$



Age vs Height

$Var(X) = 9.17$

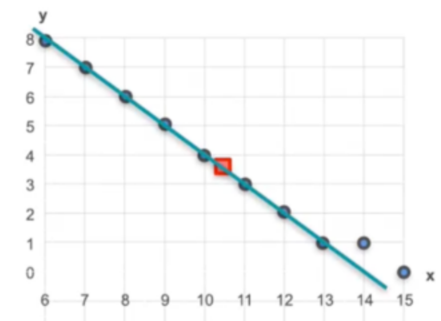
$Var(Y) = 39.56$

$\boxed{Cov(X, Y) = 17}$

When this is a positive
value it tells us that

DeepLearning.AI

Correlation Coefficient

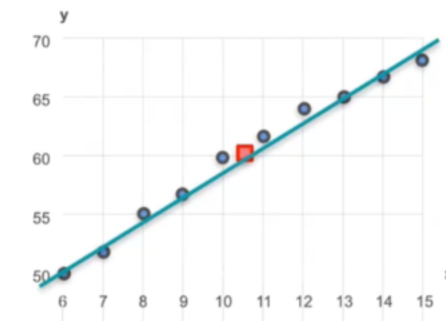


Age vs Naps per Day

$Var(X) = 9.17$

$Var(Y) = 7.57$

$Cov(X, Y) = -7.45$



Age vs Height

$Var(X) = 9.17$

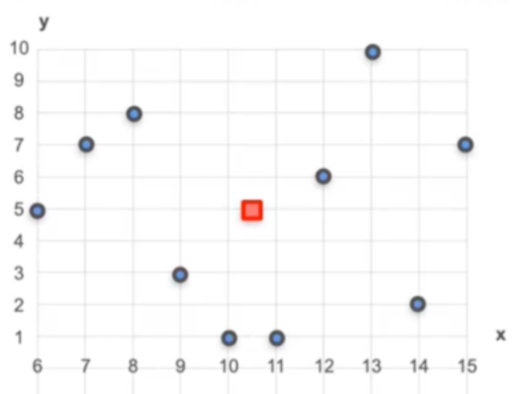
$Var(Y) = 39.56$

$Cov(X, Y) = 17$

grow however because of 17 is
bigger than 7.45 forgetting about

DeepLearning.AI

Correlation Coefficient



Age vs Grades

$Var(X) = 9.17$

$Var(Y) = 9.78$

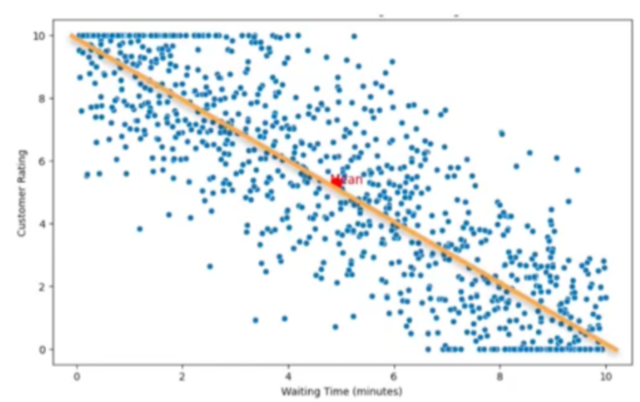
$Cov(X, Y) = 0.1$

$$\textbf{Correlation Coefficient} = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}} = \frac{0.1}{\sqrt{9.17} \cdot \sqrt{9.78}}$$

$$\approx 0.01$$

DeepLearning.AI

Correlation Coefficient



$Var(X) = 8.526$

$Var(Y) = 10.163$

$Cov(X, Y) = -7.878$

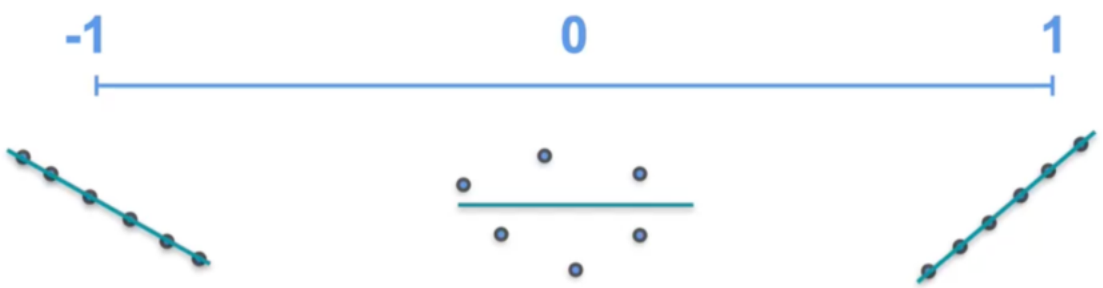
$$\textbf{Correlation Coefficient} = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}} = \frac{-7.878}{\sqrt{8.562} \cdot \sqrt{10.163}}$$

$$\approx -0.845$$

DeepLearning.AI

Correlation Coefficient

$$\textbf{Correlation Coefficient} = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$$



DeepLearning.AI

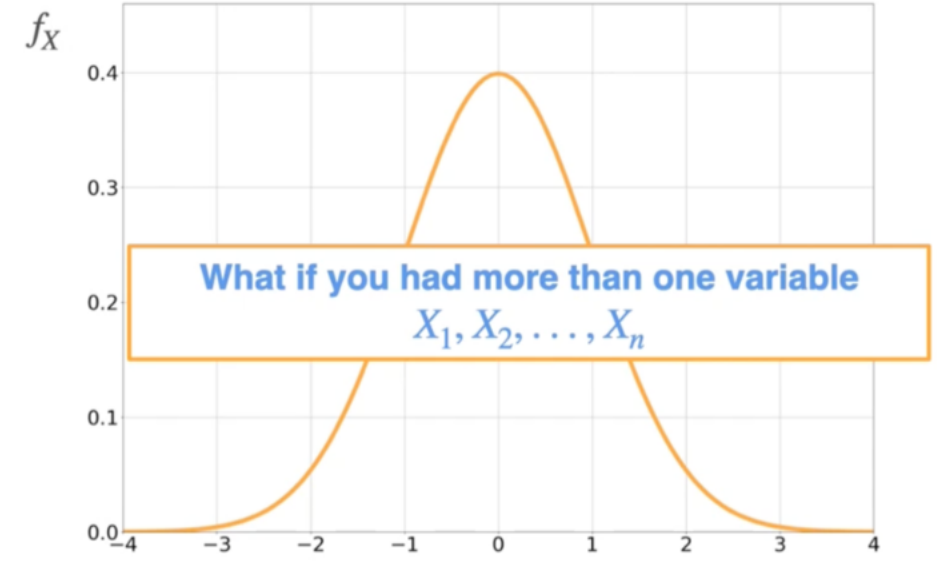
Multivariate Gaussian Distribution

For a single variable, X

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

Parameters:

- μ: center of the bell
- σ: spread of the bell



Now here's a question. What if you had more than one variable?

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

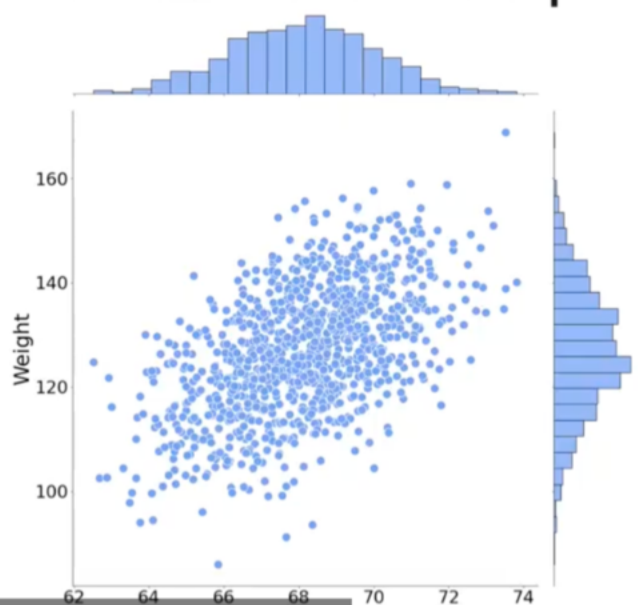
Two variables

H : Height of an adult in inches

$H \sim \mathcal{N}(\mu_H, \sigma_H)$

W : Weight of an adult in pounds

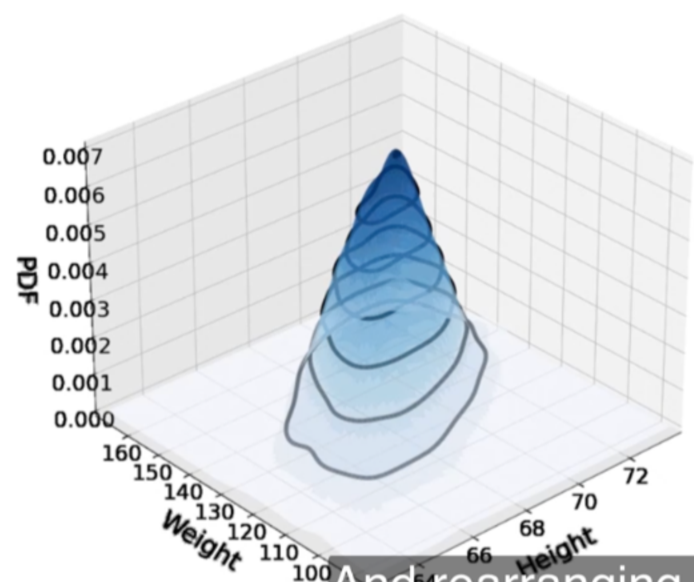
$W \sim \mathcal{N}(\mu_W, \sigma_W)$



then you can look at the margins and notice that they are both normal

DeepLearning.AI

Multivariate Gaussian Distribution: an Example



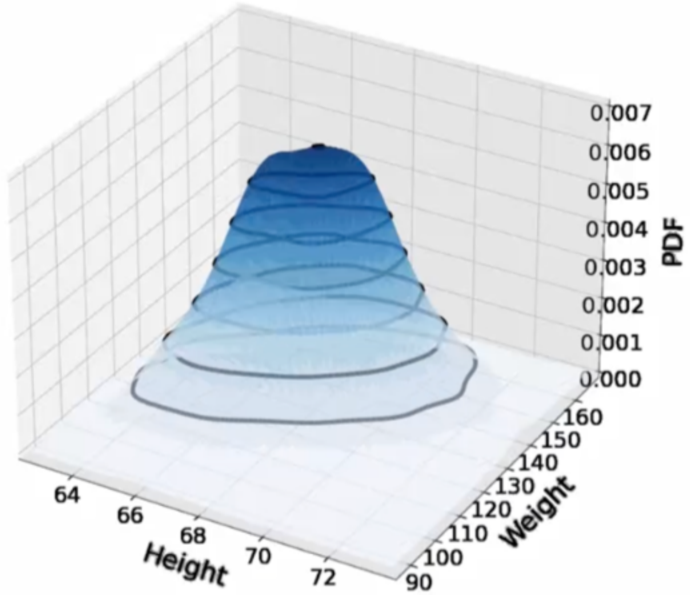
If W, H were **independent**

$$f_{HW}(h,w) = f_H(h) f_W(w)$$

And rearranging the terms a little bit, you would get this expression over here.

DeepLearning.AI

Multivariate Gaussian Distribution: an Example



If W, H were independent

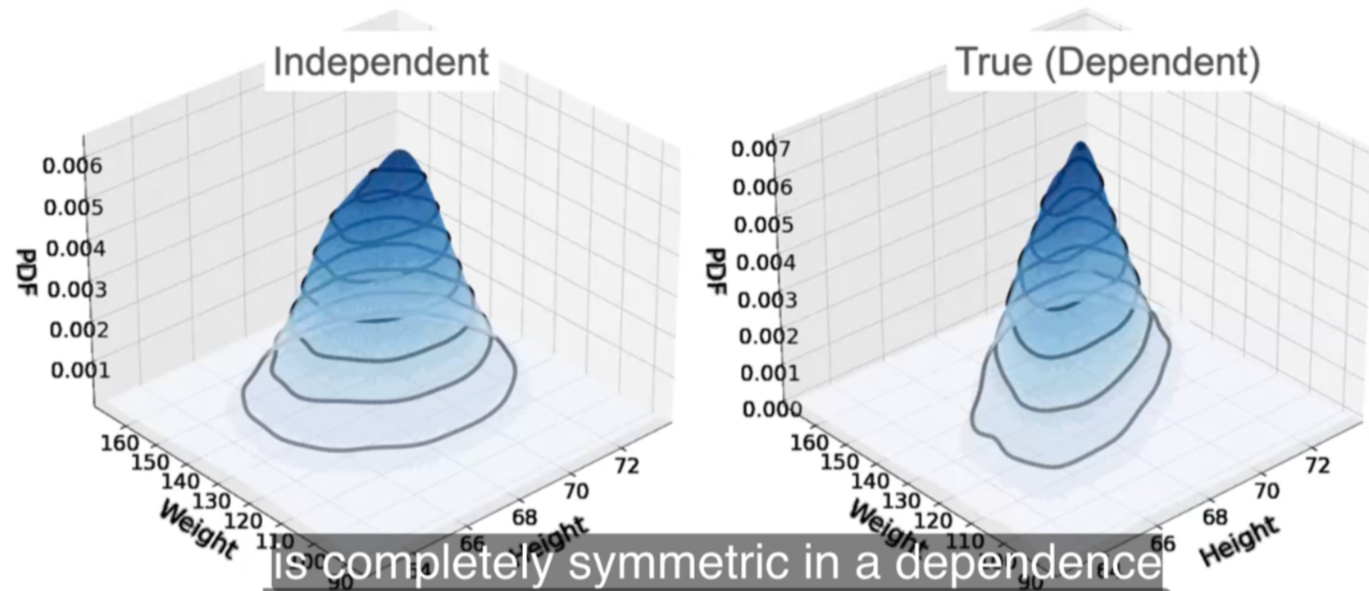
$$f_{HW}(h,w) = f_H(h)f_W(w)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_H} e^{-\frac{1}{2}\frac{(h-\mu_H)^2}{\sigma_H^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_W} e^{-\frac{1}{2}\frac{(w-\mu_W)^2}{\sigma_W^2}}$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}\right)}$$

DeepLearning.AI

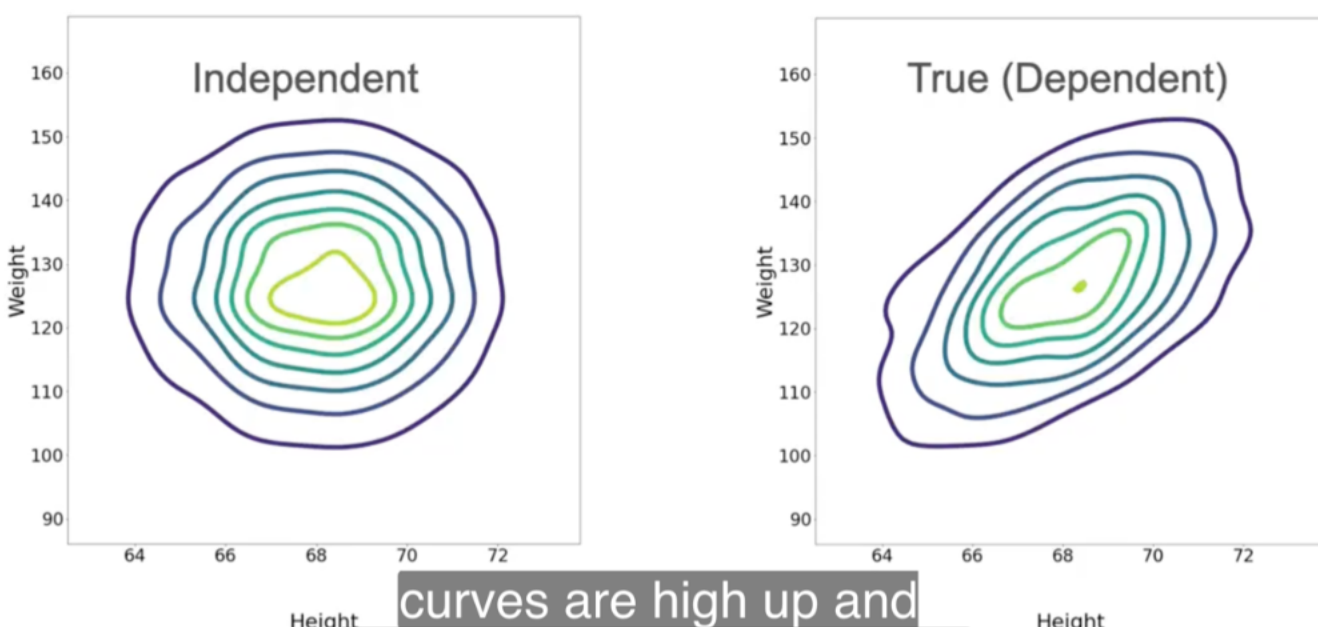
Multivariate Gaussian Distribution: an Example



is completely symmetric in a dependence
case you see that distribution is elongated

DeepLearning.AI

Multivariate Gaussian Distribution: an Example



curves are high up and
the purple curves are down.

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

If W, H were independent

$$f_{HW}(h,w) = f_H(h)f_W(w)$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\boxed{\left(\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}\right)}}$$

$$\left\| \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} \\ \frac{w-\mu_W}{\sigma_W} \end{bmatrix} \right\|_2^2 = \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} & \frac{w-\mu_W}{\sigma_W} \end{bmatrix} \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} \\ \frac{w-\mu_W}{\sigma_W} \end{bmatrix}$$

$$= ([h \quad w] - [\mu_H \quad \mu_W]) \begin{bmatrix} \frac{1}{\sigma_H^2} & 0 \\ 0 & \frac{1}{\sigma_W^2} \end{bmatrix} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)^T \begin{bmatrix} \sigma_H^2 & 0 \\ 0 & \sigma_W^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)$$

diagonal matrix with the
variances in each element

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

If W, H were independent

$$f_{HW}(h,w) = f_H(h)f_W(w)$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\boxed{\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}}\right)}$$

$$\left\| \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} \\ \frac{w-\mu_W}{\sigma_W} \end{bmatrix} \right\|_2^2 = \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} & \frac{w-\mu_W}{\sigma_W} \end{bmatrix} \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} \\ \frac{w-\mu_W}{\sigma_W} \end{bmatrix}$$

$$= ([h \quad w] - [\mu_H \quad \mu_W]) \begin{bmatrix} \frac{1}{\sigma_H^2} & 0 \\ 0 & \frac{1}{\sigma_W^2} \end{bmatrix} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)$$

Covariance matrix! (Σ)

$$= \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)^T \boxed{\begin{bmatrix} \sigma_H^2 & 0 \\ 0 & \sigma_W^2 \end{bmatrix}}^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)$$

Notice that since the variables are independent this has to be the diagonal.

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

If W, H were independent

$$f_{HW}(h,w) = f_H(h)f_W(w)$$

$$= \frac{1}{2\pi\boxed{\sigma_H\sigma_W}} e^{-\frac{1}{2}\left(\boxed{\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}}\right)}$$

$\det(\Sigma)^{1/2}$

$$\left\| \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} \\ \frac{w-\mu_W}{\sigma_W} \end{bmatrix} \right\|_2^2 = \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} & \frac{w-\mu_W}{\sigma_W} \end{bmatrix} \begin{bmatrix} \frac{h-\mu_H}{\sigma_H} \\ \frac{w-\mu_W}{\sigma_W} \end{bmatrix}$$

$$= ([h \quad w] - [\mu_H \quad \mu_W]) \begin{bmatrix} \frac{1}{\sigma_H^2} & 0 \\ 0 & \frac{1}{\sigma_W^2} \end{bmatrix} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)$$

Covariance matrix! (Σ)

$$= \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)^T \boxed{\begin{bmatrix} \sigma_H^2 & 0 \\ 0 & \sigma_W^2 \end{bmatrix}}^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \boxed{\begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix}} \right)$$

μ

the determinant of the covariance matrix.

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

If W, H were independent

$$f_{HW}(h,w) = f_H(h)f_W(w)$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}\right)}$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} \exp\left(-\frac{1}{2} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right)^T \begin{bmatrix} \sigma_H^2 & 0 \\ 0 & \sigma_W^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \begin{bmatrix} \mu_H \\ \mu_W \end{bmatrix} \right) \right)$$

Remember that this is valid only if w and h were independent.

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

Dependent case:

$$f_{HW}(h,w) = f_H(h)f_W(w)$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}\right)}$$

$$= \frac{1}{2\pi \det \Sigma^{1/2}} \exp\left(-\frac{1}{2} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \mu \right)^T \Sigma^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \mu \right) \right)$$

Well, of course, it's no longer true that the joint

DeepLearning.AI

Multivariate Gaussian Distribution: an Example

1

Dependent case:

$$f_{HW}(h, w) = \cancel{f_H(h) f_W(w)}$$

$$\cancel{= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}\right)}}$$

$$\Sigma = \begin{bmatrix} \sigma_H^2 & Cov(H, W) \\ Cov(H, W) & \sigma_W^2 \end{bmatrix}$$

$$= \frac{1}{2\pi \det \Sigma^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} h \\ w \end{bmatrix} - \mu\right)^T \Sigma^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \mu\right)\right)$$

and now the off-diagonal amounts correspond to the covariances between variables.

DeepLearning.AI

2

Multivariate Gaussian Distribution: an Example

Dependent case:

$$f_{HW}(h, w) = \cancel{f_H(h) f_W(w)}$$

$$\cancel{= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\frac{(h-\mu_H)^2}{\sigma_H^2} + \frac{(w-\mu_W)^2}{\sigma_W^2}\right)}}$$

$$\Sigma = \begin{bmatrix} \sigma_H^2 & Cov(H, W) \\ Cov(H, W) & \sigma_W^2 \end{bmatrix}$$

$$= \frac{1}{2\pi \det \Sigma^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} h \\ w \end{bmatrix} - \mu\right)^T \Sigma^{-1} \left(\begin{bmatrix} h \\ w \end{bmatrix} - \mu\right)\right)$$

and now the off-diagonal amounts correspond to the covariances between variables.

DeepLearning.AI

Multivariate Gaussian Distribution: General Definition

3

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

$$f_X(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \left|\Sigma\right|^{\frac{1}{2}}}$$

covariance matrix
spread of the bell

$|\Sigma|$ determinant of the covariance matrix

The bigger the determinant the bigger the spread.

DeepLearning.AI

4

Multivariate Gaussian Distribution: General Definition

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

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; \cdots \;\; x_n]^T$$

Mean vector

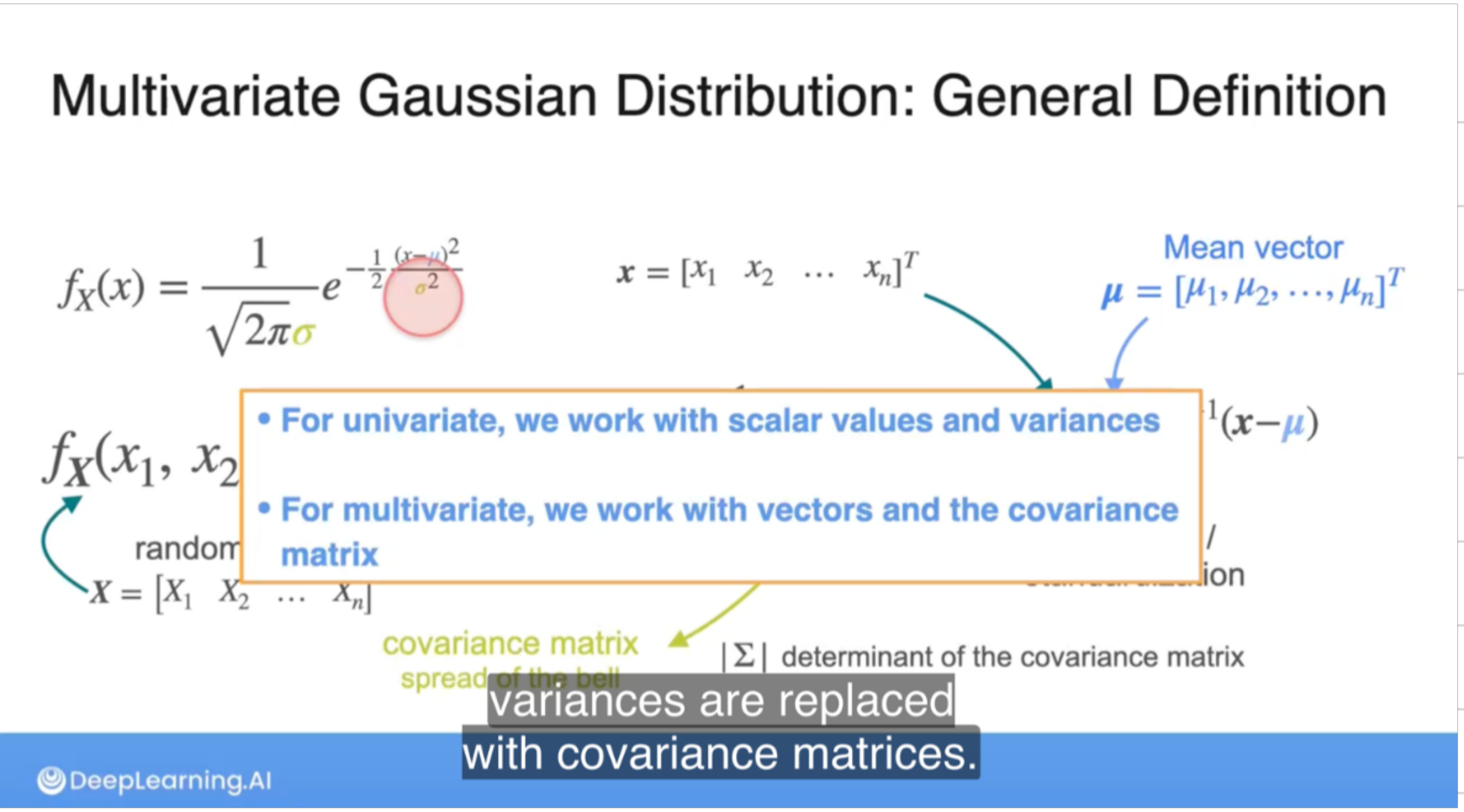
$$\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_n]^T$$

$$f_X(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \left|\Sigma\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$$

covariance matrix
spread of the bell

$|\Sigma|$ determinant of the covariance matrix

DeepLearning.AI



Gaussian Distribution Multivariate :

μ and σ (Σ) are vectors.

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

Where $x = [x_1, x_2 \, . \, . \, x_n]^T$

$\mu = [\mu_1, \mu_2 \, . \, . \, . \, \mu_n]^T$

So $|\Sigma|$ determinant of the Covariance Matrix

Source: Coursera Probability & Statistics for Data Science & Machine Learning!