Day-59, Jan 28, 2025 (Magh 15, 2081 B.S.) [Jan 28]

Sum of Gaussians: an Example:

R: Total Response time of a Computer System

T: processing time

$T \sim N(10, 2^2)$

Independent

L: Network Latency

$L \sim N(5, 1^2)$

$$\boxed{R = T + L}$$

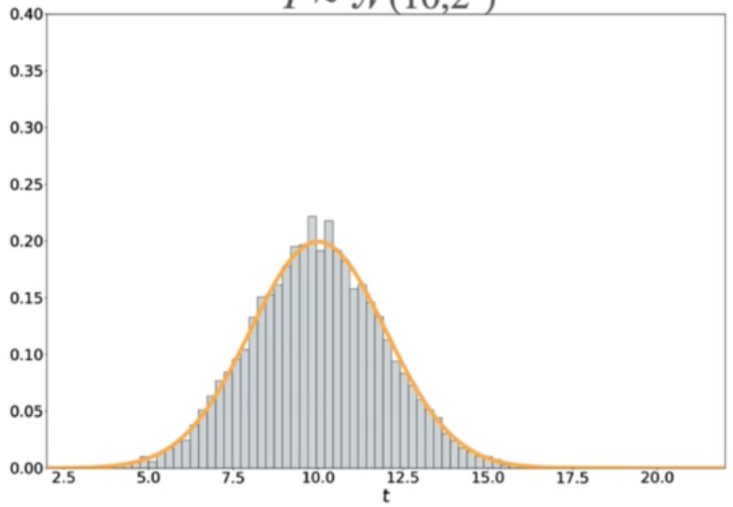
Parameters of these Gaussian = ? they are $\mu_R = E[R]$

$= E[T] + E[L]$

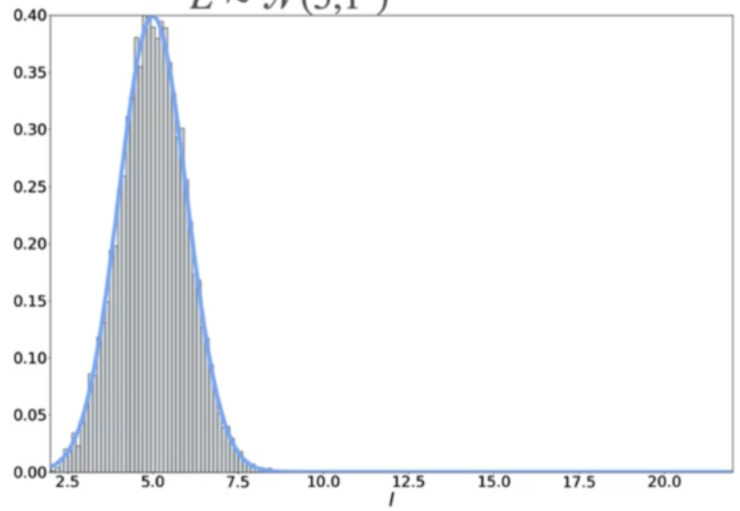
Sum of Gaussians

Sample each variable 10000 times

$T \sim \mathcal{N}(10, 2^2)$



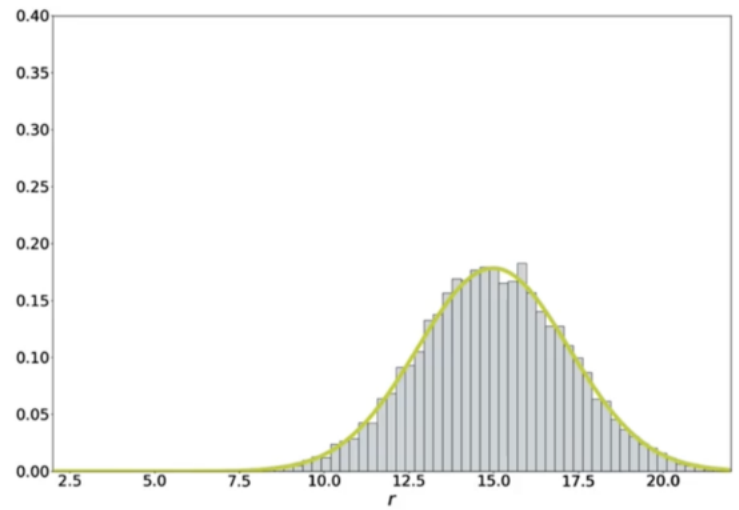
$L \sim \mathcal{N}(5, 1^2)$



DeepLearning.AI

Sum of Gaussians

$R = T + L$



R is still Gaussian!

The histogram of $R = T + L$ is Gaussian.

DeepLearning.AI

R is still Gaussian!

Expectation is linear

$$\mu_R = E[R] = E[T] + E[L]$$

$$\mu_R = E[R]$$

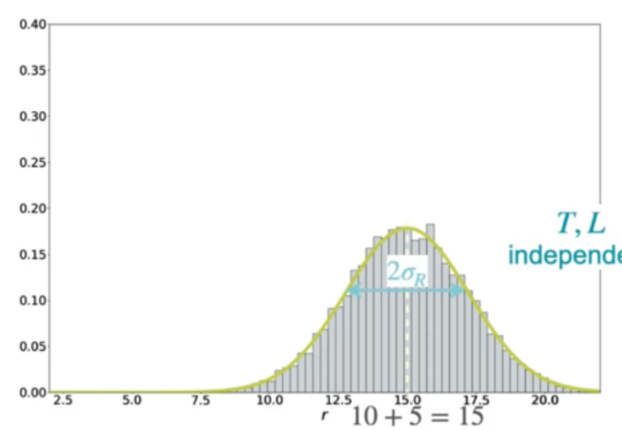
$$\Rightarrow E[T + L]$$

$$\Rightarrow E[T] + E[L]$$

$$\sigma_R^2 = Var(R)$$

Sum of Gaussians

$R = T + L$



R is still Gaussian!

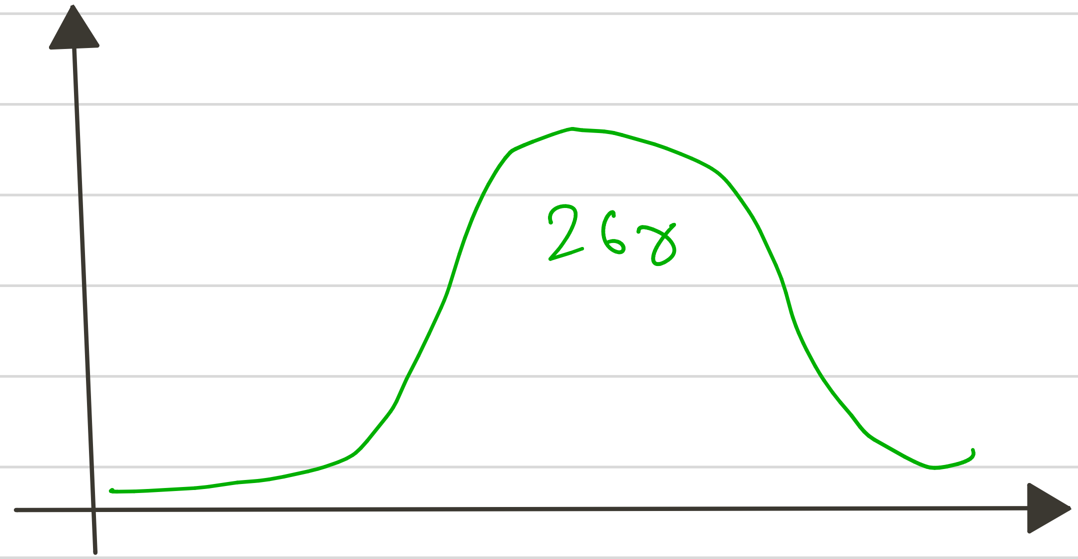
Expectation is linear

$$\mu_R = \mathbb{E}[R] = \mathbb{E}[T + L] = \mathbb{E}[T] + \mathbb{E}[L]$$
$$= \mu_T + \mu_L = 10 + 5 = 15$$

T, L independent

$$\sigma_R^2 = Var(R) \quad Var(T + L)$$
$$= Var(T) + Var(L) = \sigma_T^2 + \sigma_L^2$$
$$= 4 + 1 = 5$$

So, $R = (T+L) \sim N(10+5, 4+1)$



So, In general: $W = aX + bY$

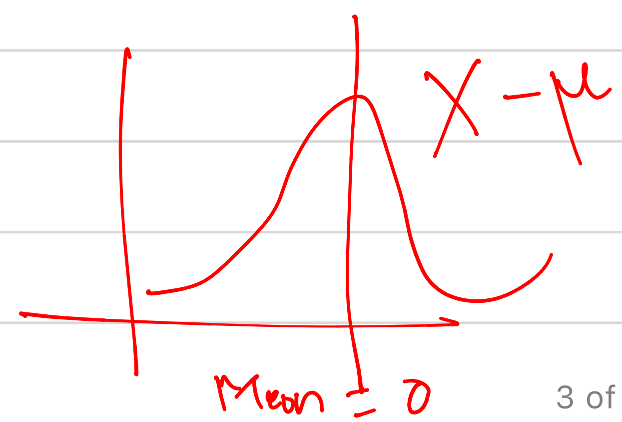
Independent $\begin{cases} X \sim N(\mu_X, \sigma_X^2) \\ Y \sim N(\mu_Y, \sigma_Y^2) \end{cases}$

$$\rightarrow W \sim N(a\mu_x + b\mu_y,\ a^2\sigma_x^2 + b^2\sigma_y^2)$$

Next is Standardizing a Distribution?

Everything is Nicer when the Mean is 0.

$X \rightarrow X - \mu$



μ = Mean with some $|x|$ we $X \to X - \mu$

$Y = \mu$

So, $E[X + Y] = E[X] + E[Y]$

$\Rightarrow E[X] - E[\mu]$

$\Rightarrow E[X] - \mu$

So, $E[X + Y] = \Rightarrow 0$

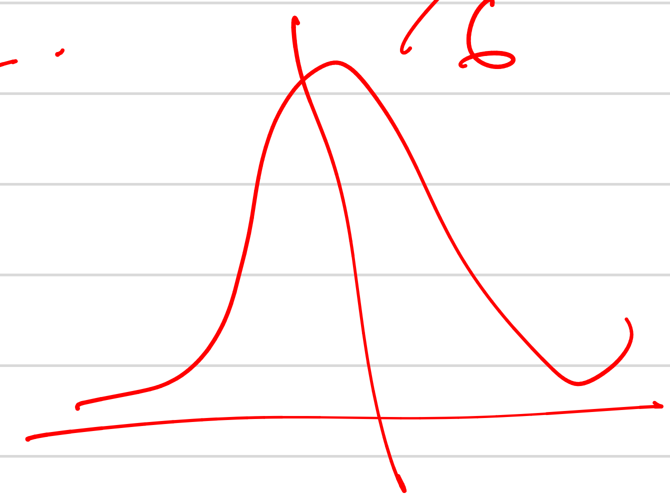
Everything is nicer where the $\sigma = 1$.

Mean = 0

Std deviation = 1

$X \to X/\sigma$

X/σ



Why we use standardizing a distribution?

→ Comparability between different datasets

→ Simplification of statistical analysis

→ Improved performance of Machine Learning Models.

→ All of the above.

Firstly it transformation datasets into a standard scale, making it easier to compare between different datasets.

Secondly, it simplifies statistical analysis, particularly when using techniques that assumes a standard Normal Distribution.

Finally, the standardizing features in machine learning can improve the convergence rate of optimization algorithms and prevent some features from dominating others, leading to improved model performance.

Proof: on Everything is Nicer when the σ is 1.

$$Var(cX) = E[(cX)^2] - E[cX]^2$$

$$\Rightarrow E[c^2X^2] - (cE[X])^2$$

$$\Rightarrow c^2E[X^2] - c^2E[X]^2$$

$$\Rightarrow c^2(E[X^2] - E[X]^2)$$

$$\Rightarrow c^2 Var(X)$$

We have,

$$\mathrm{Var}\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2}\,\mathrm{Var}(X)$$

Squaring both sides we get,

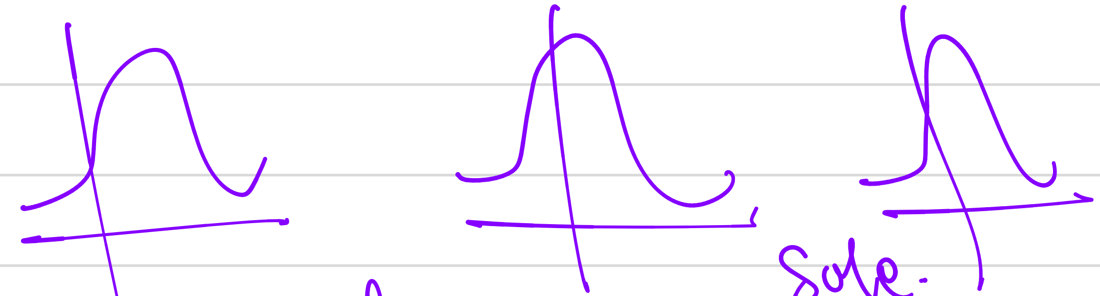
$$\mathrm{Std}\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma}\,\mathrm{Std}(X)$$

$$\Rightarrow \frac{\sigma}{\sigma}$$

$$\Rightarrow 1.$$

Standardize?

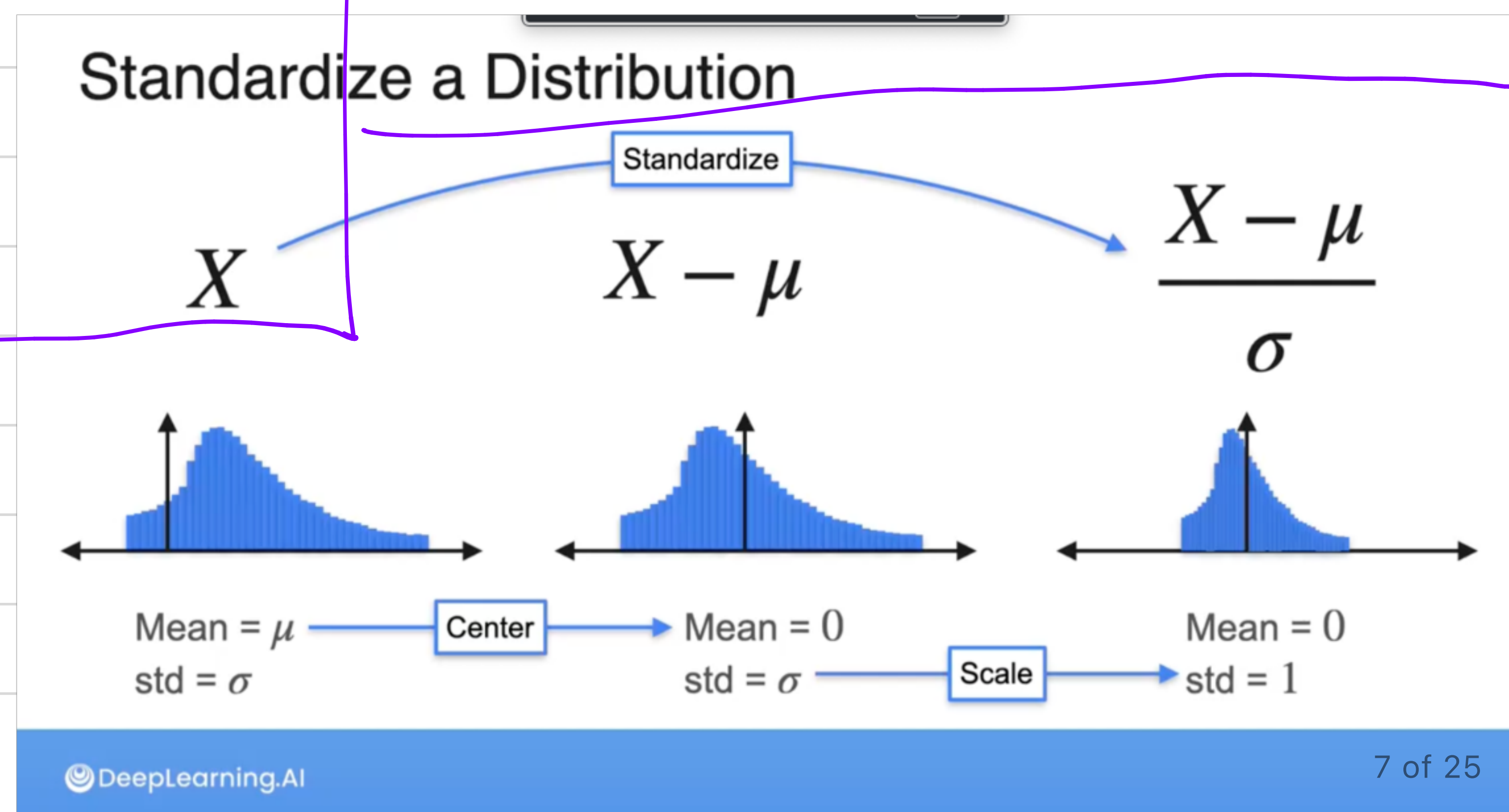
$X \xrightarrow{\text{Standardize}} X-\mu \rightarrow \frac{X-\mu}{\sigma}$



$\mu.$ σ $\xrightarrow{\text{Center}}$ $\mu = 0$ σ $\xrightarrow{\text{Scale}}$ $\mu = 0$ $\sigma = 1$

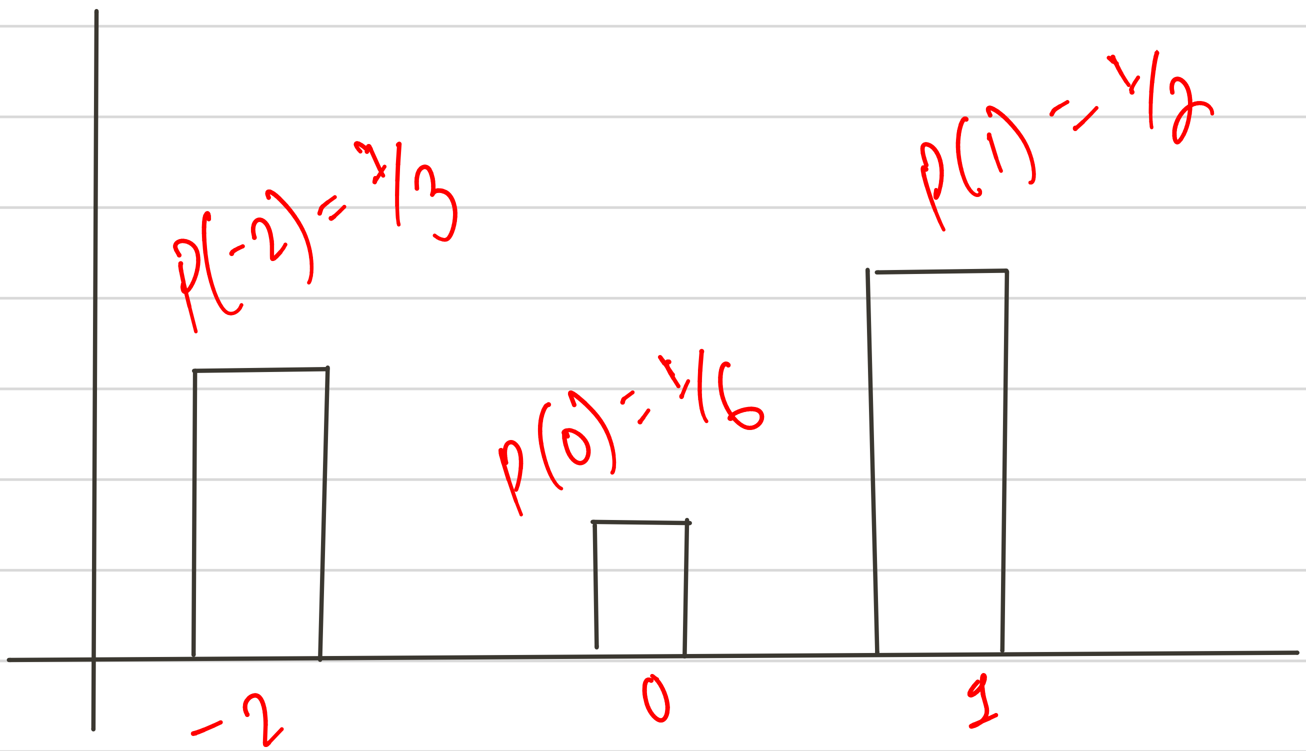
If $\mu = -2$ and $\sigma = 4$ then

$$X = \frac{X-(-2)}{4} \Rightarrow \frac{X+2}{4}$$



Skewness and Kurtosis: Moments of Distribution:

Expectation.



Random variable X.

Histogram are Probability

$\Rightarrow$ then.

$$E[X] = \frac{1}{3}(-2) + \frac{1}{6}(0) + \frac{1}{2}(1)$$

$$E[X^2] = \frac{1}{3}(-2)^2 + \frac{1}{6}(0)^2 + \frac{1}{2}(1)^2$$

$$E[X^3] = \frac{1}{3}(-2)^3 + \frac{1}{2}(0)^3 + \frac{1}{2}(1)^3$$

$$\vdots$$

$$E[X^k] = \frac{1}{3}(-2)^k + \frac{1}{6}(0)^k + \frac{1}{2}(1)^k$$

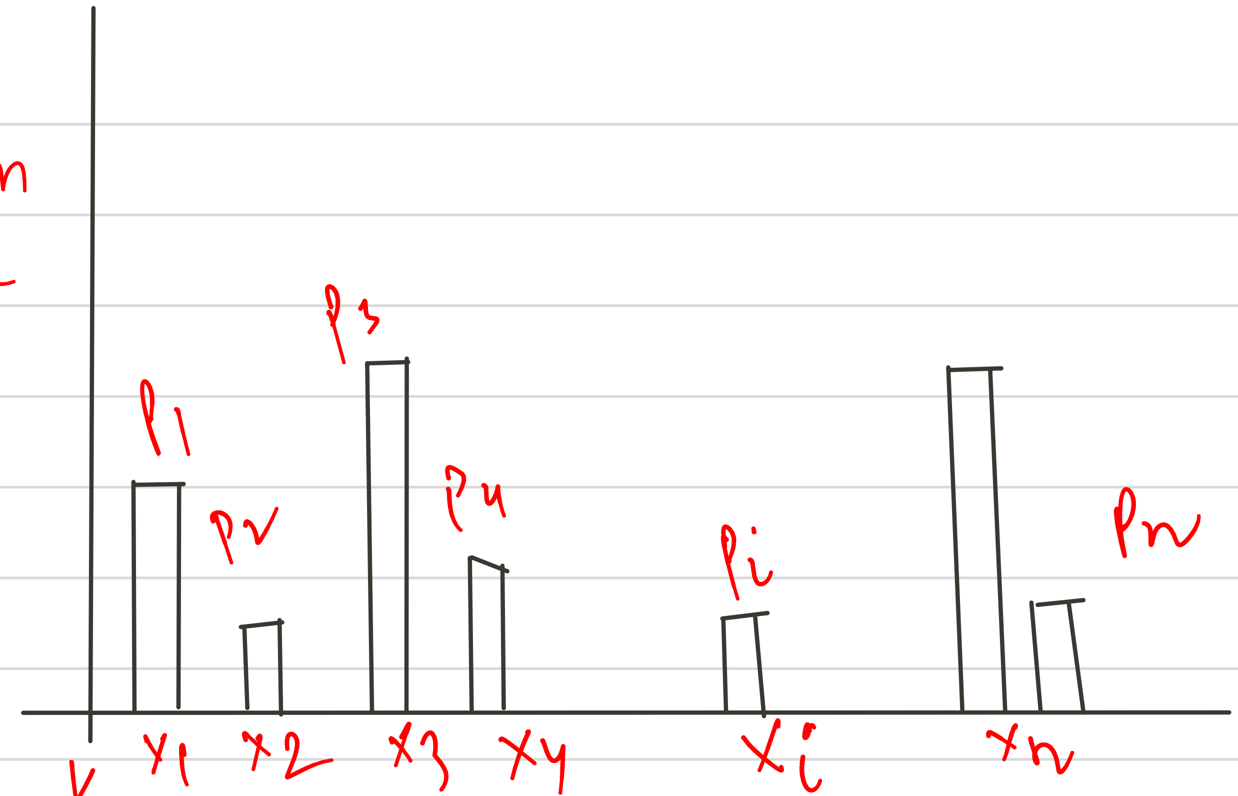
Moments of Distribution:

$$E[X] = p_1 x_1 + p_2 x_2 + p_3 x_3 + \cdots + p_n x_n$$

$$E[X^2] = p_1 x_1^2 + p_2 x_2^2 + \cdots + p_n x_n^2$$

$$\vdots$$

$$E[X^k] = p_1 x_1^k + p_2 x_2^k + \cdots + p_n x_n^k$$



Lottery vs Insurance

Ticket: \$1
Jackpot: \$100

Lottery

Cost: \$1
Crash Reparation: \$100

Car insurance

You **win** \$99 with 1% probability

You **lose** \$1 with 99% probability

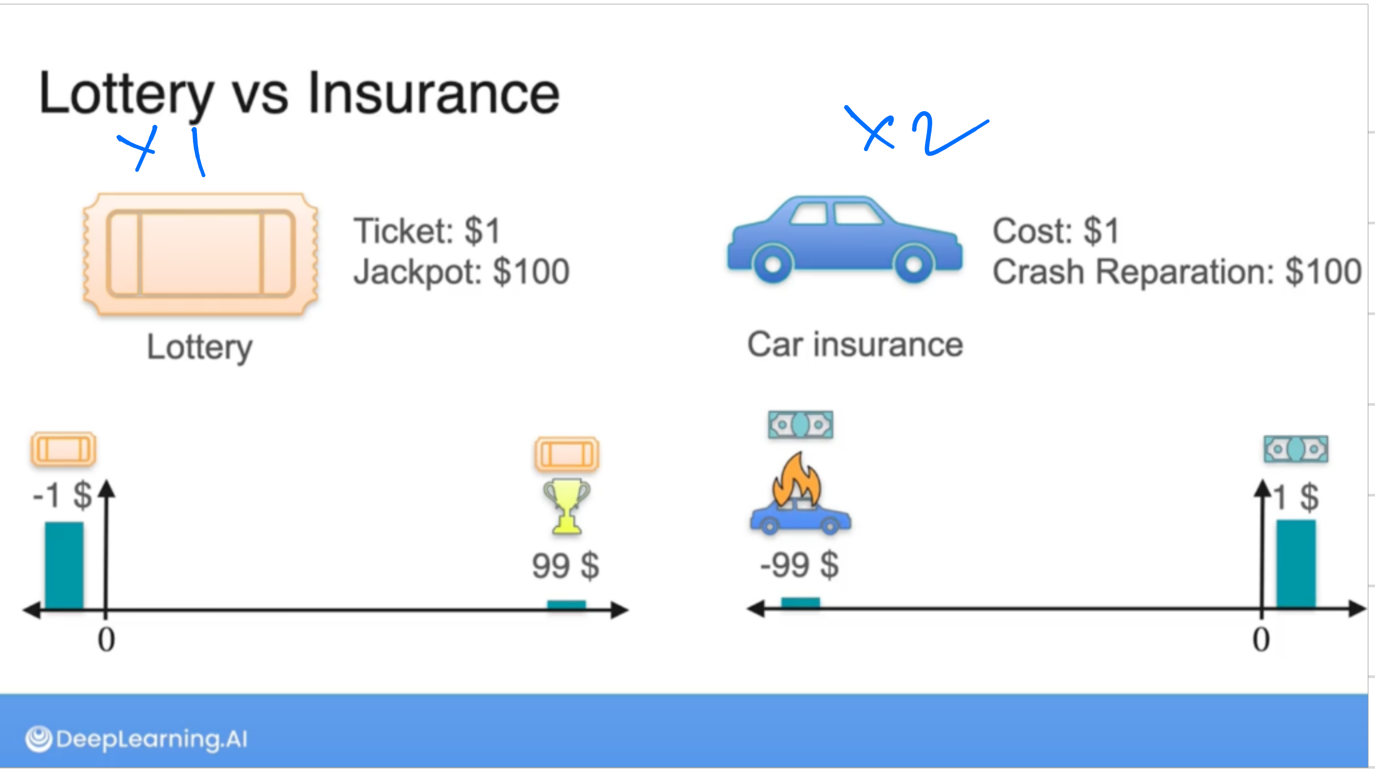
You **win** \$1 with 99% probability

You **lose** \$99 with 1% probability

So the lottery and Insurance are just opposite to each other

Having same variance of £.

Let's see if it is detected by $E[X]$ and $Var(X)$.



Calculation:

Expectation = $E[x]$.

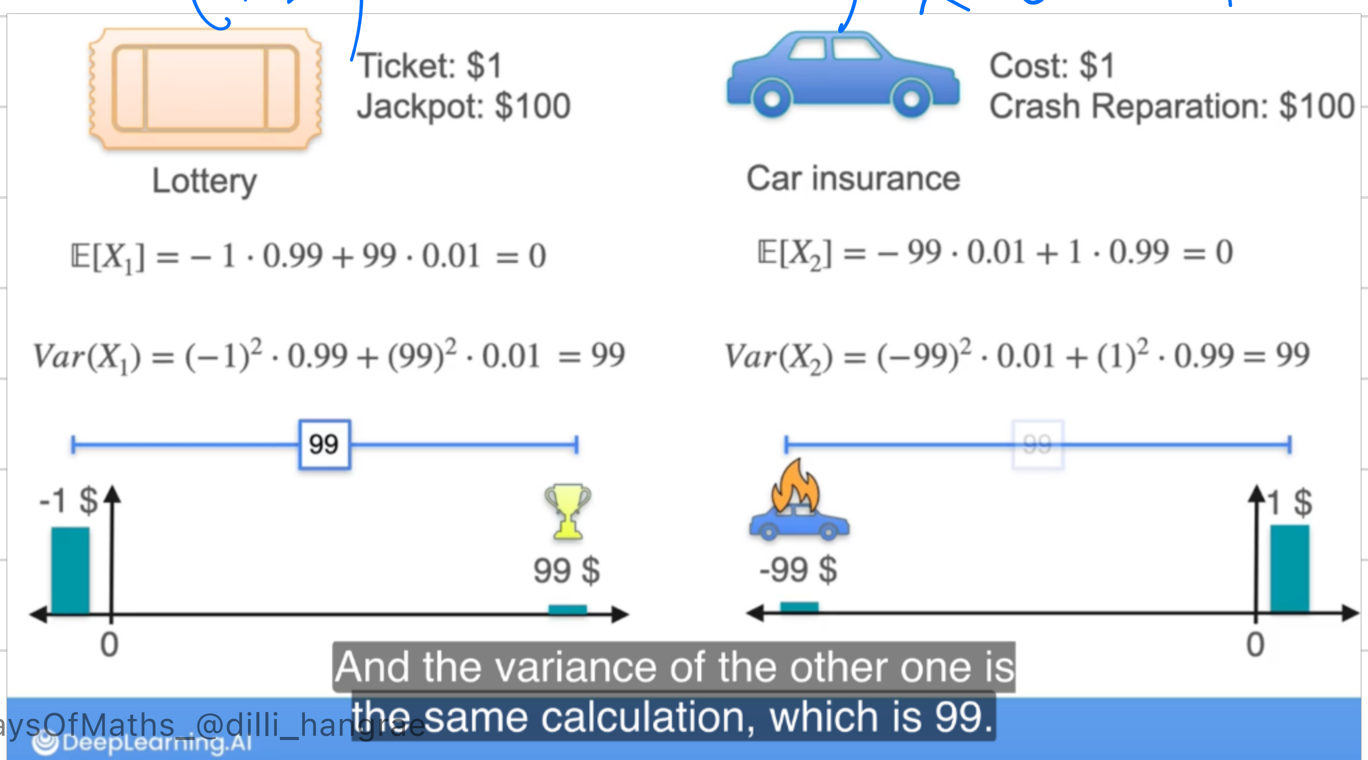
$$E[X_1] = -1 \cdot 0.99 + 99 \times 0.01 \Rightarrow 0$$

$$E[X_2] \Rightarrow -99 \times 0.01 + 1 \times 0.99 \Rightarrow 0$$

Variance = ?

$$Var(X_1) = (-1)^2 \cdot 0.99 + (99)^2 \times 0.01 \Rightarrow 99$$

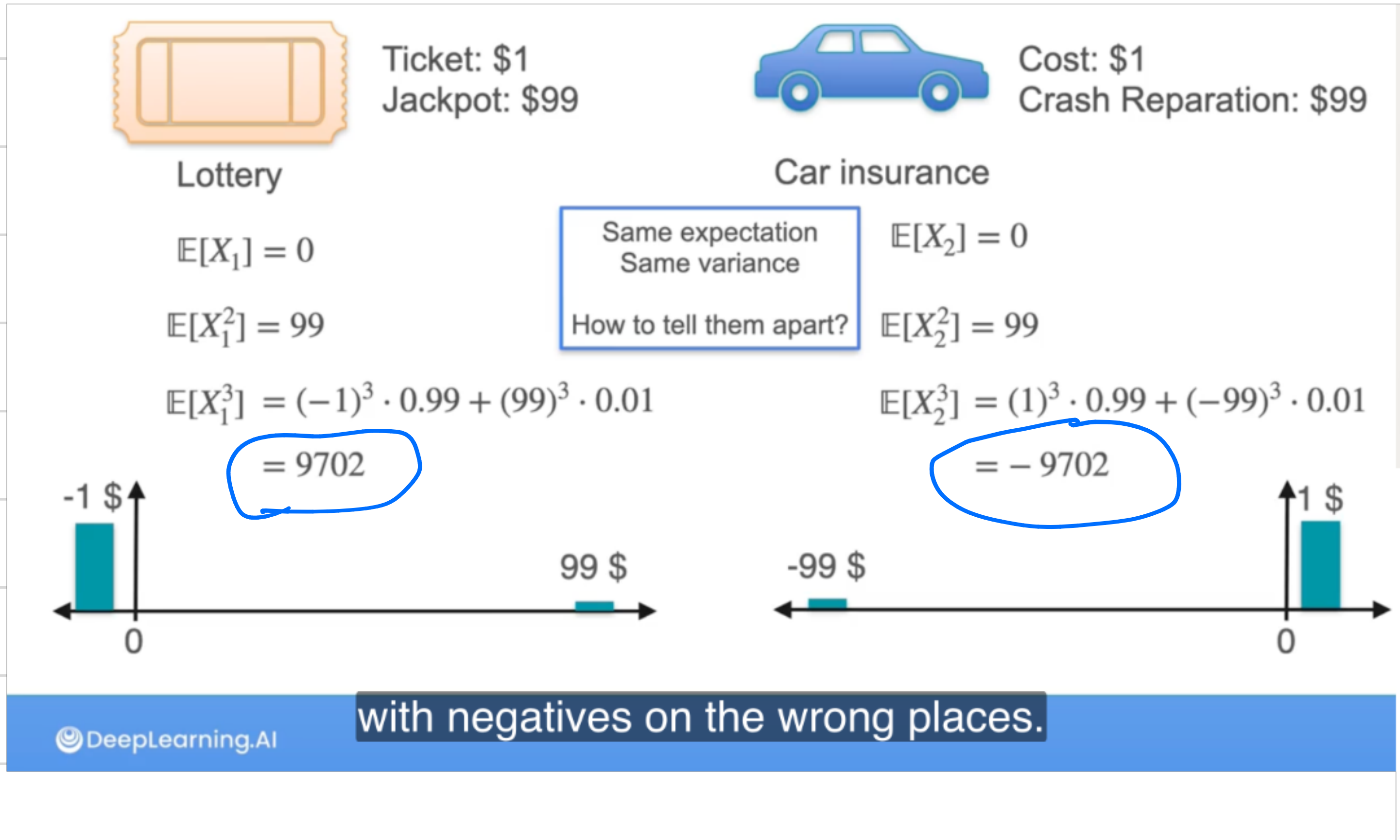
$$Var(X_2) = (-99)^2 \times 0.01 + (1)^2 \times 0.99 \Rightarrow 99$$



But, How to tell them Apart?..

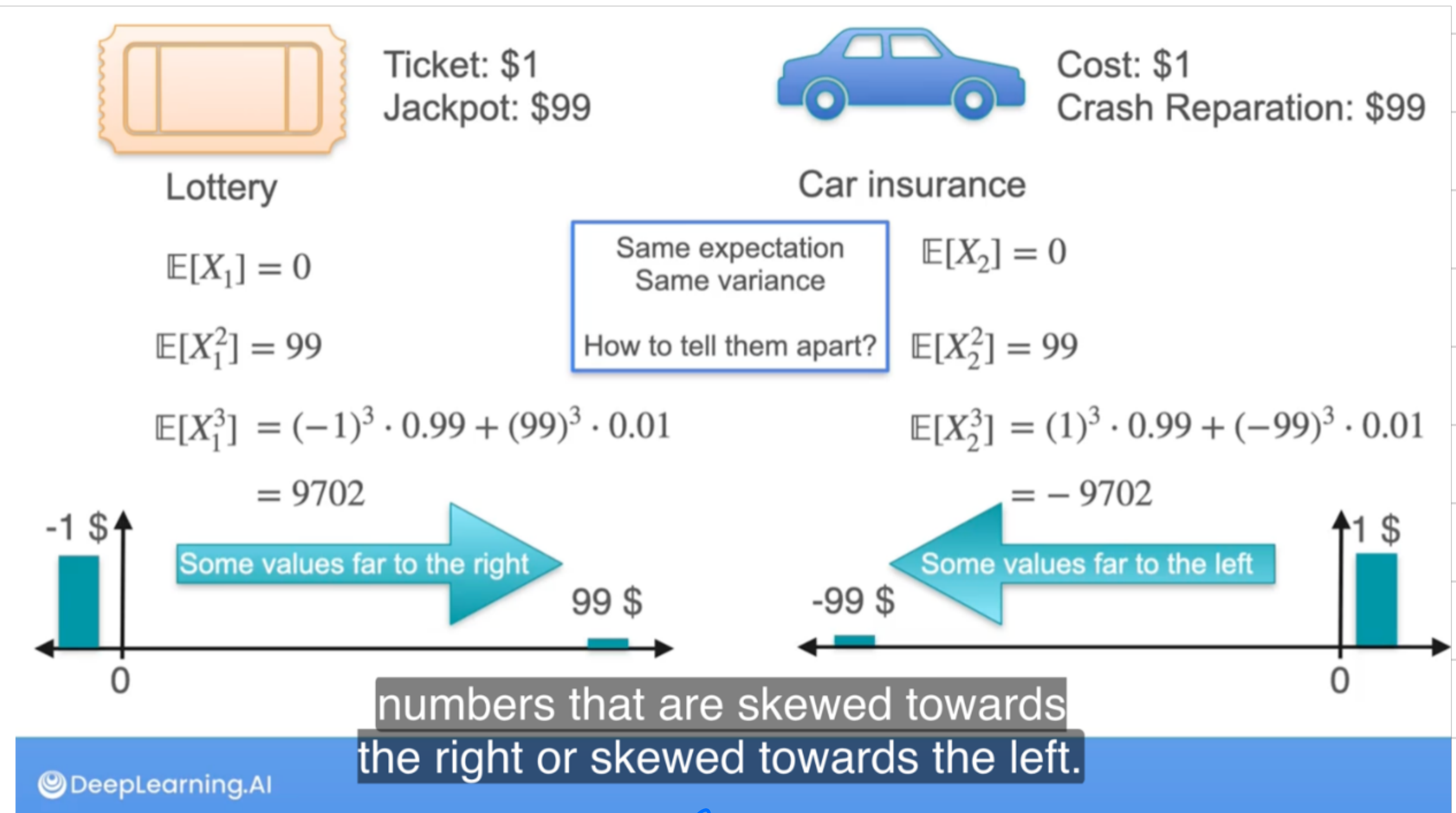
We saw two moments X_1 and X_2

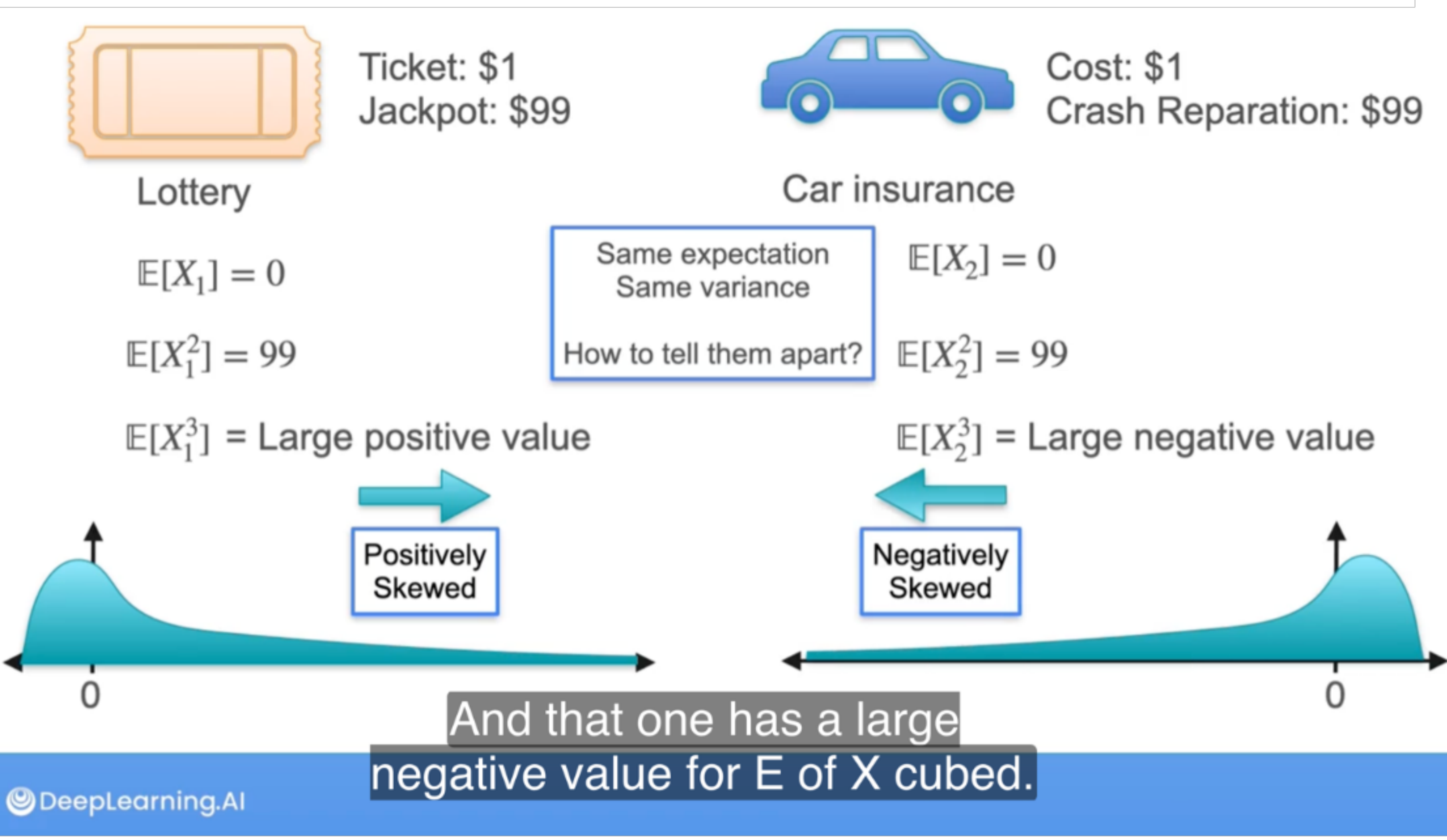
So why don't we use third moment?



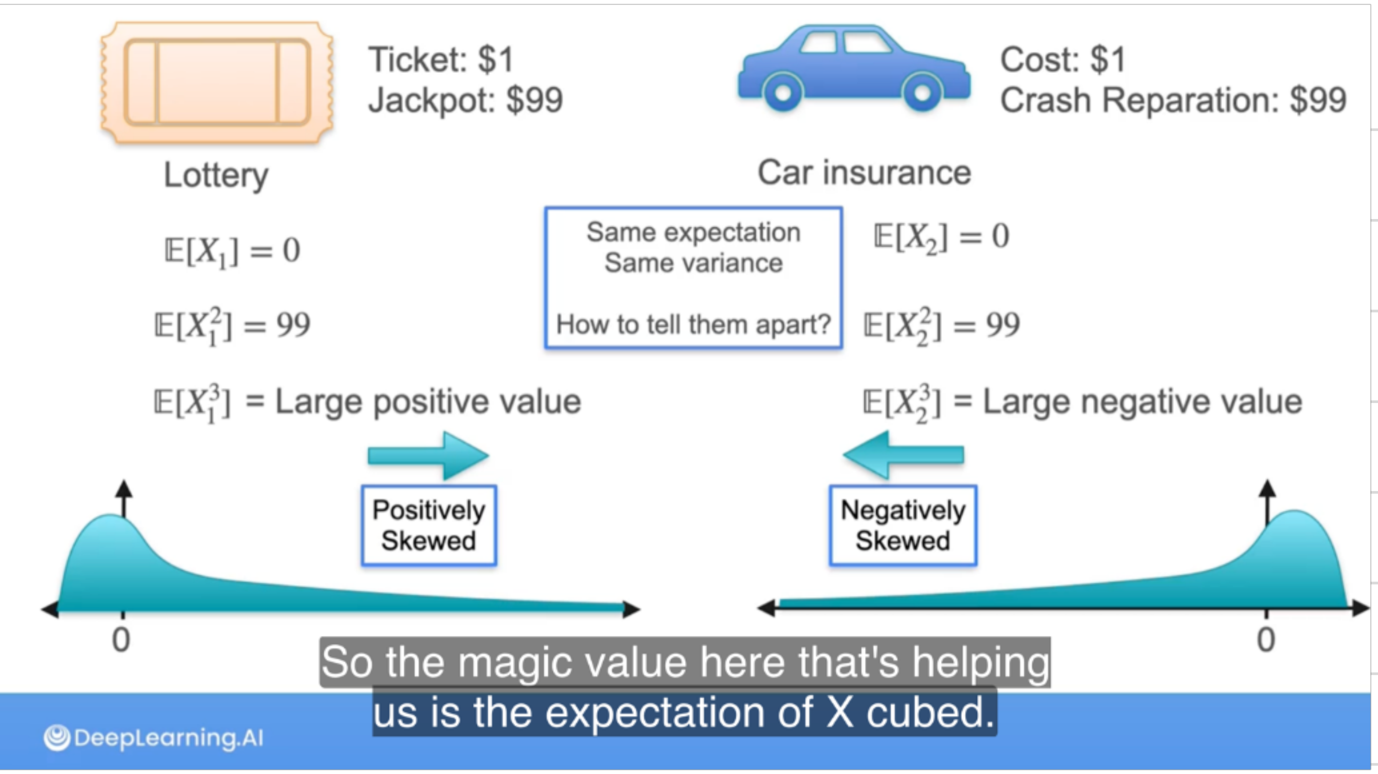
Largely skewed +ve vs largely skewed -ve.

$-E[X_1^3]$ and $+E[X_2^3]$





So, the cube^3 of the variable detects the numbers that are skewed towards the right or skewed towards the left

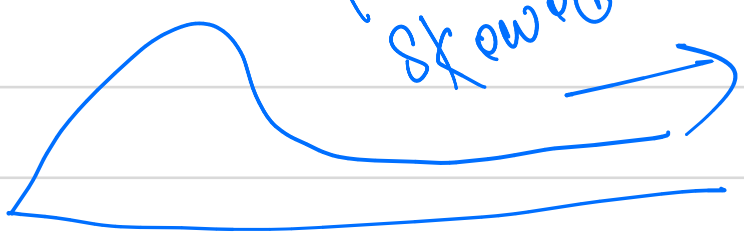


$E[X]^3$ so the skewness is the expected value of the centered and standardized distribution.

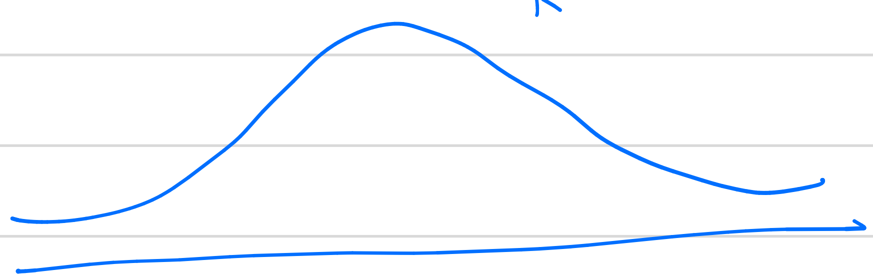
$\mathbb{E}[X^3]$ Almost Need to standardize.

$$\text{Skewness} = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

+vely skewed



Not skewed



+vely skewed

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] > 0 \qquad E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = 0 \qquad E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] < 0$$

Skewness and Kurtosis: Kurtosis

Example:

Game 1

probability $\frac{1}{2}$: you win 1 dollar

" $\frac{1}{2}$: you lose 1 dollar

Game 2

probability $\frac{100}{202}$: you win 10 cents

" $\frac{100}{202}$: you lose 10 cents

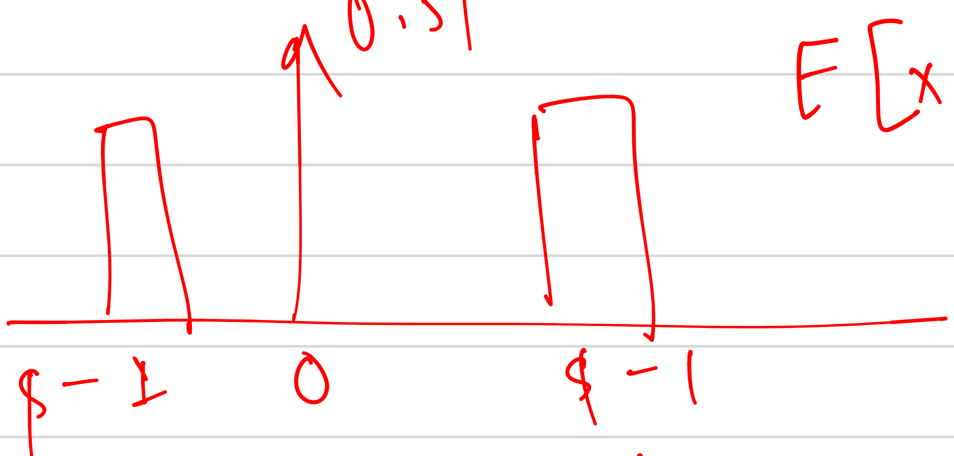
" $\frac{1}{202}$: you win 10 dollars

" $\frac{1}{202}$: you lose 10 dollars

$E[X] = ?$ $\quad$ $Var(X) = ?$

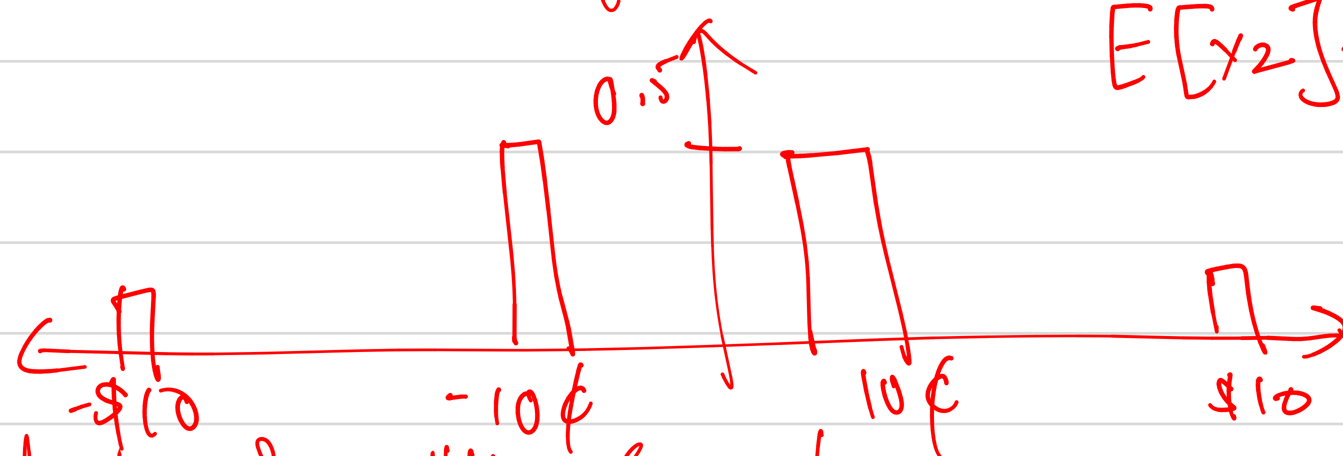
So, the plots are

Game 1.



$E[X_1] = 0$

Game 2



$E[X_2] = 0$

Both has same variance But how? We'll show it -

$$E[X_1^2] = \frac{1}{2}(-1)^2 + \frac{1}{2}(1)^2$$

$$E[X_1^2] \Rightarrow 1$$

$$Var(X_1) = 1$$

$$Std(X_1) = 1$$

$$Skew(X_1) = 0$$

$$E[X_2^2] = \frac{100}{202}(-0.1)^2 + \frac{100}{202}(0.1)^2 + \frac{1}{202}(-10)^2 + \frac{1}{202}(10)^2$$

$$\Rightarrow \frac{100}{202} \cdot \frac{1}{100} + \frac{100}{202} \cdot \frac{1}{100} + \frac{1}{202} \cdot 100 + \frac{1}{202} \cdot 100$$

$$E[X_2^2] \Rightarrow 1$$

$$Var(X_2) = 1$$

$$Std(X_2) = 1$$

Q. How do we tell them apart? $Skew(X_2) = 0$

Since there Moments are

	Moment	
$E[X_1] = 0$	1st	$E[X_2] = 0$
$E[X_1^2] = 1$	2nd	$E[X_2^2] = 1$
$E[X_1^3] = 0$	3rd	$E[X_2^3] = 0$

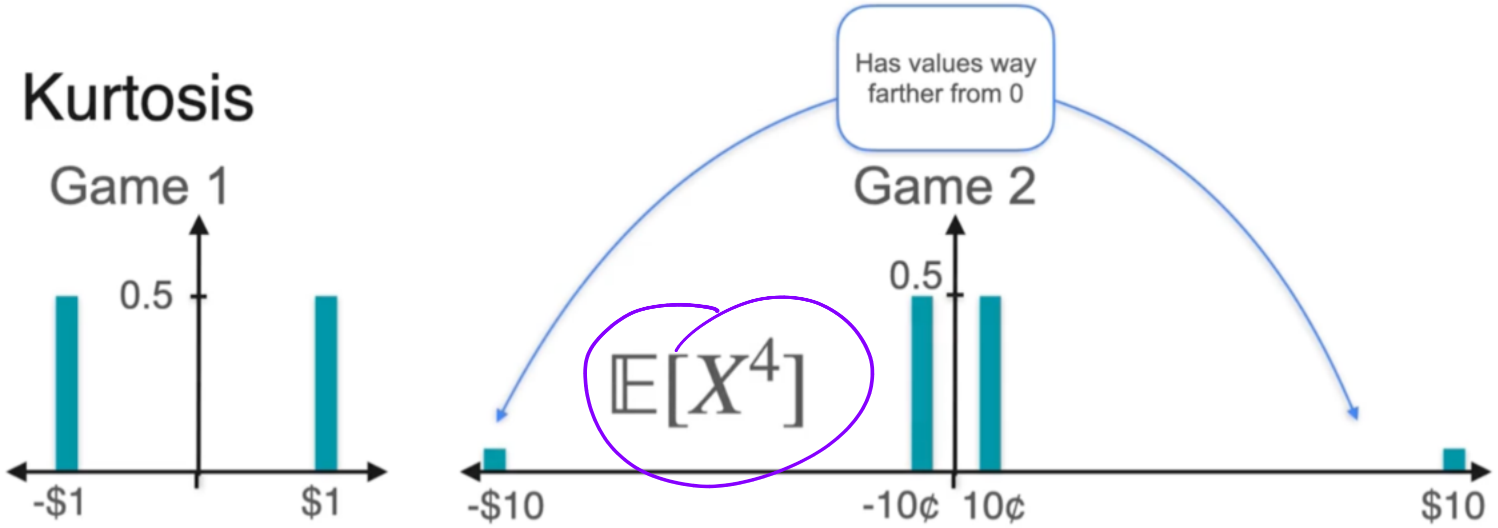
$E[X_1] = 0$

$Var(X_1) = 1$

$Skew(X_1) = 0$

Q. Who'll help to know difference?

Kurtosis



$E[X_1]=0$	$E[X_1]=0$	$E[X_2]=0$	$E[X_2]=0$
$Var(X_1)=1$	$E[X_1^2]=1$	$Var(X_2)=1$	$E[X_2^2]=1$
$Skew(X_1)=0$	$E[X_1^3]=0$	$Skew(X_2)=0$	$E[X_2^3]=0$

DeepLearning.AI

What about the fourth moment?

Fourth Moment $\mathbb{E}[x^4]$

$$\mathbb{E}[x_1^4] = \frac{1}{2}(-1)^4 + \frac{1}{2}(1)^4$$

$$\therefore \mathbb{E}[x_1^4] = 1.$$

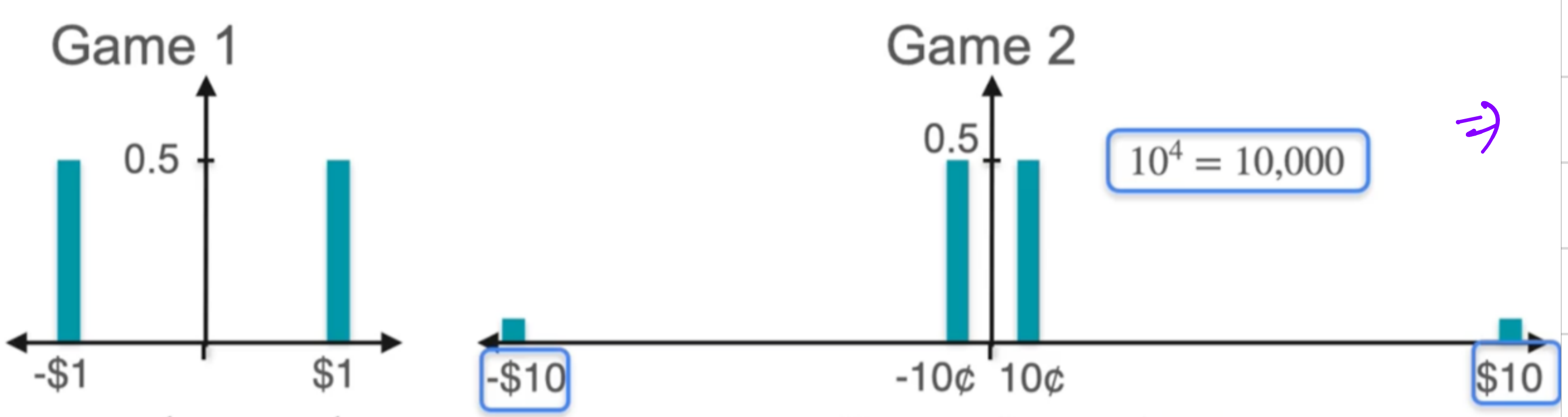
$$\mathbb{E}[x_2^4] \Rightarrow \frac{100}{202}(-0.1)^4 + \frac{100}{202}(0.1)^4 + \frac{1}{202}(-10)^4 + \frac{1}{202}(10)^4$$

$$\Rightarrow \mathbb{E}[x_2^4] \Rightarrow 99.01$$

Now, let's standardize this

Fourth Moment. $\mathbb{E}[x_n^4]$ (Kurtosis.)

Kurtosis



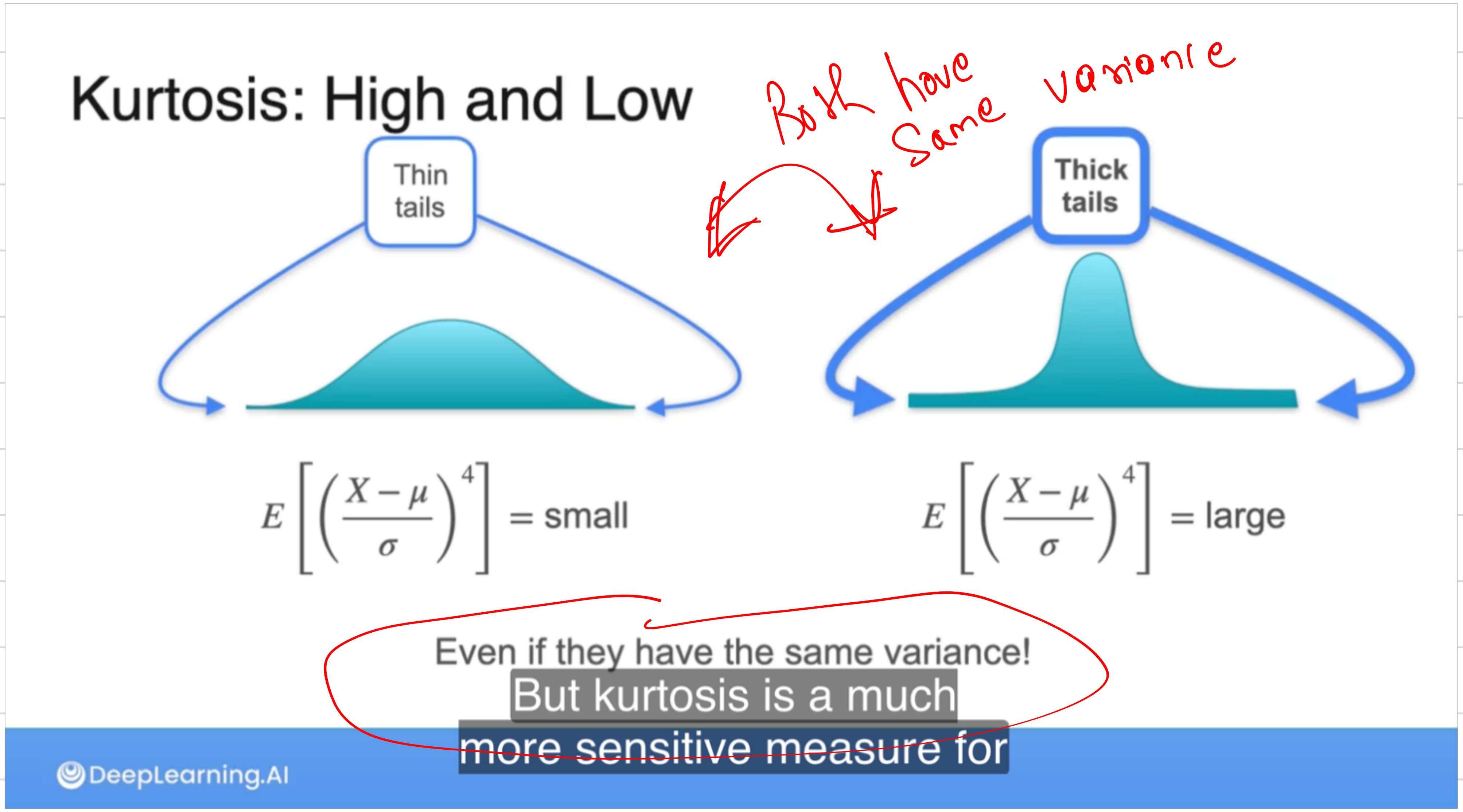
$$\mathbb{E}[X_1^4] = \frac{1}{2}(-1)^4 + \frac{1}{2}(1)^4 = 1$$

$$\mathbb{E}[X_2^4] = \frac{100}{202}(-0.1)^4 + \frac{100}{202}(0.1)^4 + \frac{1}{202}\boxed{(-10)^4} + \frac{1}{202}\boxed{(10)^4} = 99.01$$

DeepLearning.AI

So, that is called kurtosis.

Kurtosis (Fourth Moment) $\Rightarrow E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$



Both have same variance

So, Kurtosis is much more sensitive measure for the thickness of the tails of the distribution.

So, the methods to tell distributions apart expected value

① Variance or Std. Deviation

② Skewness

③ Kurtosis

$E[X_1]$ $E[X_1^2]$ $E[X_1^3]$ $E[X_1^4]$

Several ways to visulize the Data

Quantiles: Example

Newspaper advertisement

1

Newspaper Advertisement (X)			
18.3	18.4	23.2	51.2
35.2	29.7	75	8.7
65.9	14.2	54.7	25.9

For simplicity, let's consider only twelve samples from the newspaper ad budget.

DeepLearning.AI

Quantiles: Example

What is the median here?

2

The point that splits your data in half

$$\text{Median} = \frac{25.9 + 29.7}{2} = 27.8$$

50% quantile

Second quartile

Newspaper Advertisement (X)			
8.7	14.2	18.3	18.4
23.2	25.9	29.7	35.2
51.2	54.7	65.9	75

That is 27.8.

DeepLearning.AI

Quantiles: Example

3

What about the point that leaves 1/4 of your data to the left and 3/4 to the right?

$$q_{0.25} = Q1 = \frac{18.3 + 18.4}{2} = 18.35$$

25% quantile

First quartile

Newspaper Advertisement (X)			
8.7	14.2	18.3	18.4
23.2	25.9	29.7	35.2
51.2	54.7	65.9	75

into one quarter three quarters.

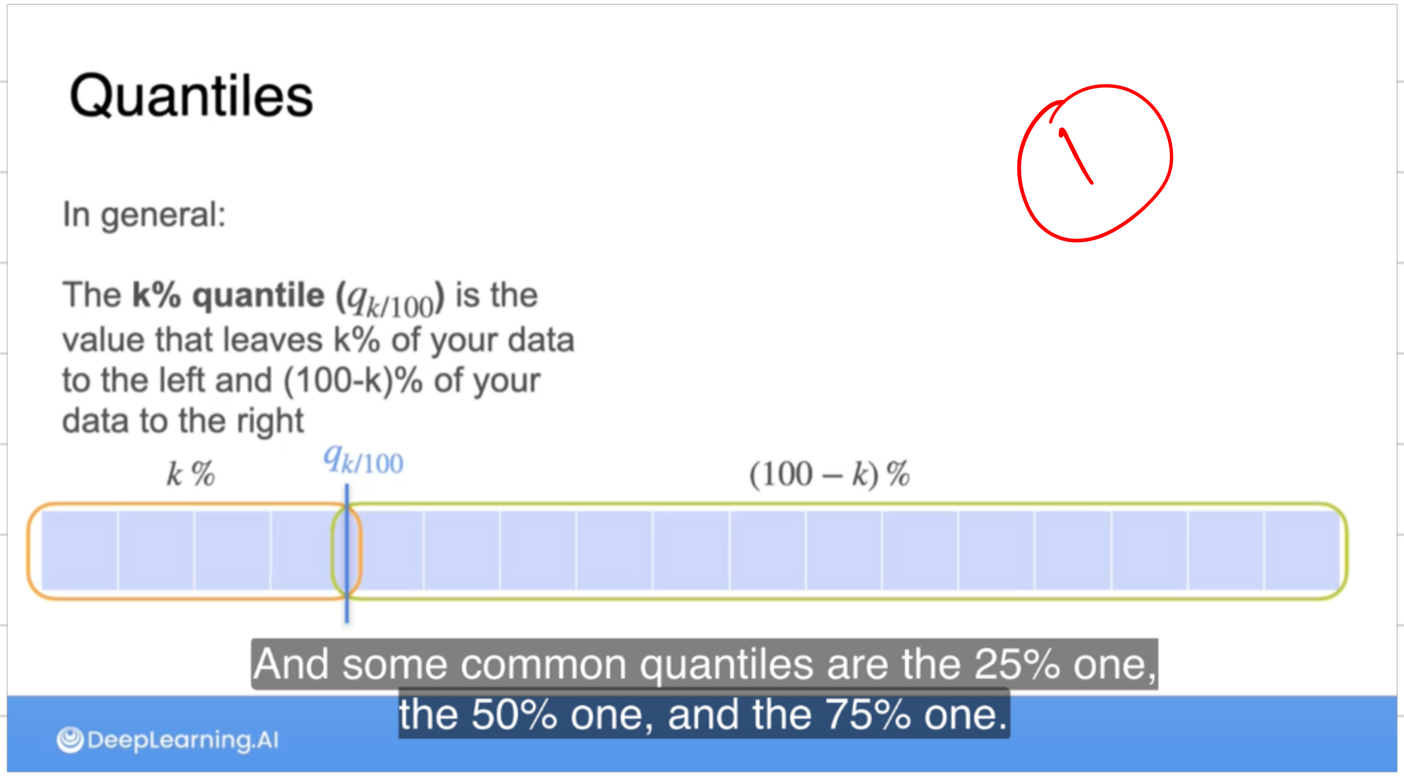
100DaysOfMaths_@dilli...

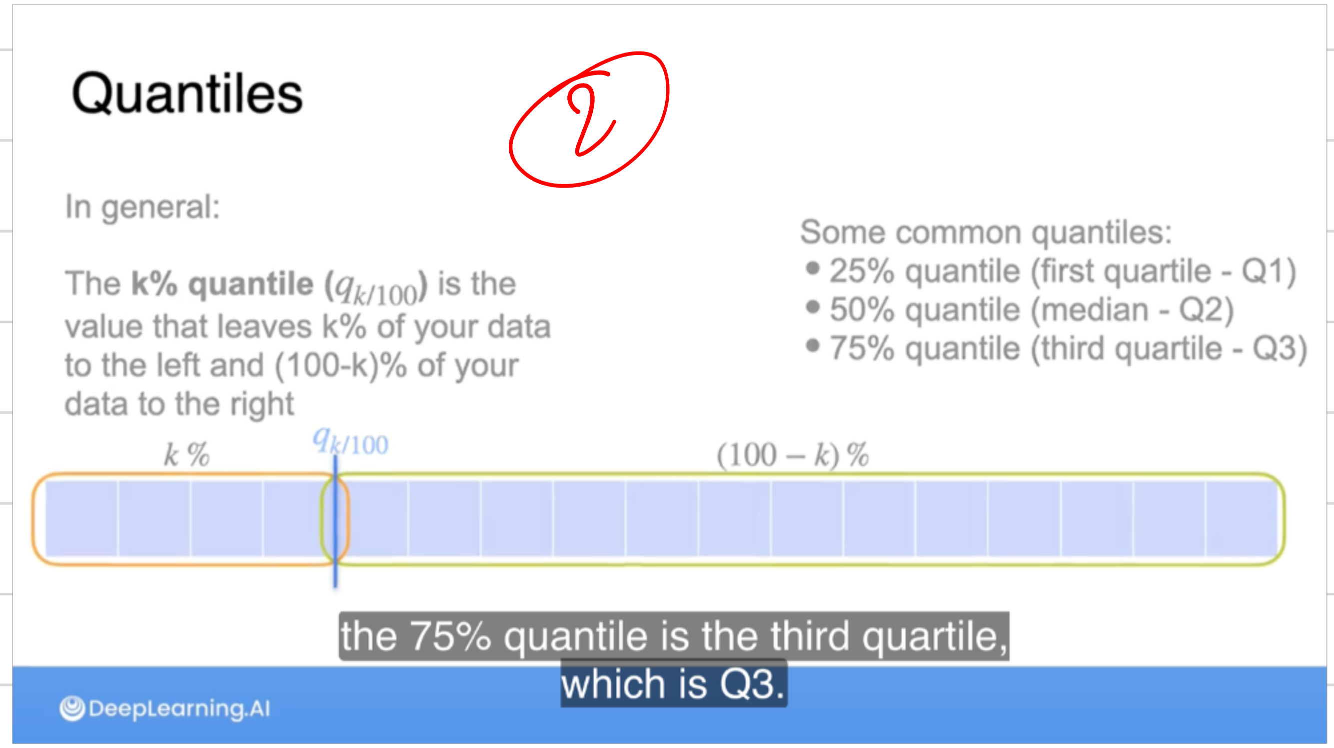
DeepLearning.AI

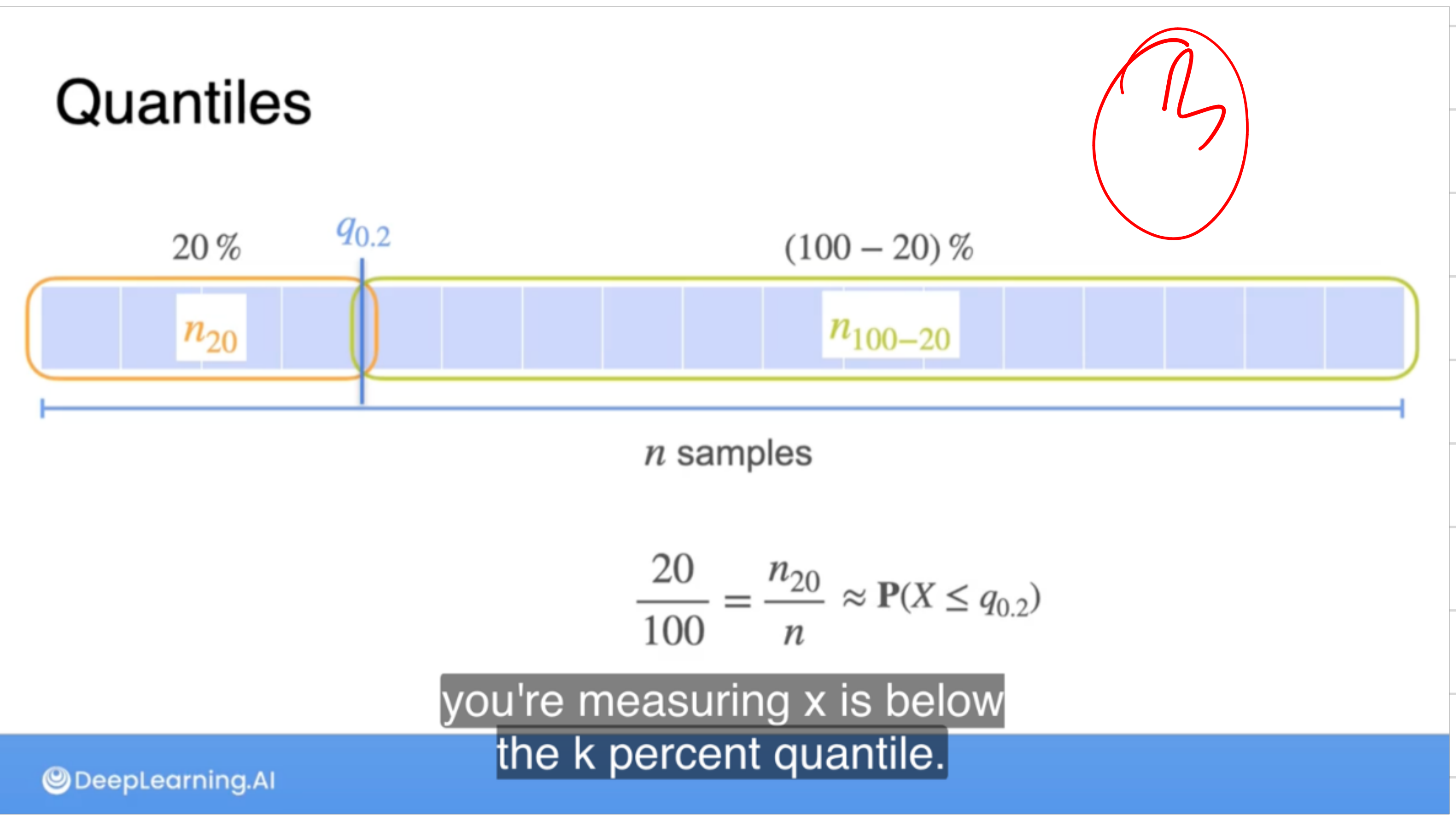
Median divides data into two parts.

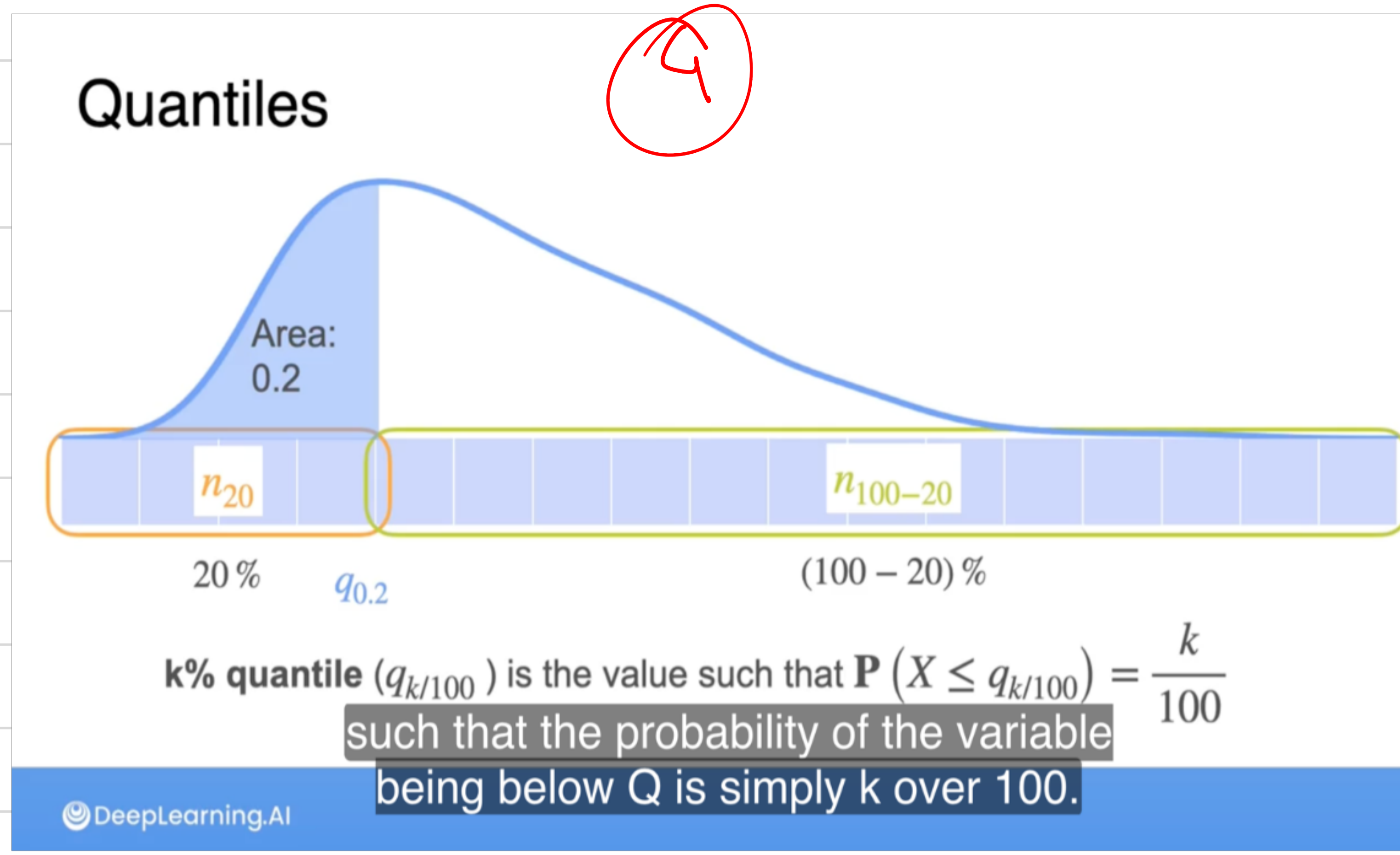
Q_1, Q_2, Q_3

Q_1 = Median. 50%.


Quantiles
1
In general:
The **k% quantile** ($q_{k/100}$) is the value that leaves k% of your data to the left and (100-k)% of your data to the right
$k\,\%$
$q_{k/100}$
$(100-k)\,\%$
And some common quantiles are the 25% one,
the 50% one, and the 75% one.
DeepLearning.AI


Quantiles
2
In general:
The **k% quantile** ($q_{k/100}$) is the value that leaves k% of your data to the left and (100-k)% of your data to the right
Some common quantiles:
• 25% quantile (first quartile - Q1)
• 50% quantile (median - Q2)
• 75% quantile (third quartile - Q3)
$k\,\%$
$q_{k/100}$
$(100-k)\,\%$
the 75% quantile is the third quartile,
which is Q3.
DeepLearning.AI


Quantiles
3
$20\,\%$
$q_{0.2}$
$(100-20)\,\%$
n_{20}
n_{100-20}
n samples
$\frac{20}{100} = \frac{n_{20}}{n} \approx \mathbf{P}(X \le q_{0.2})$
you're measuring x is below
the k percent quantile.
DeepLearning.AI


Quantiles
4
Area:
0.2
n_{20}
n_{100-20}
$20\,\%$
$q_{0.2}$
$(100-20)\,\%$
k% quantile ($q_{k/100}$) is the value such that $\mathbf{P}\left(X \le q_{k/100}\right) = \frac{k}{100}$
such that the probability of the variable
being below Q is simply k over 100.
DeepLearning.AI

Visualizing Data: Box: plots

Box-Plots

1

Newspaper Advertisement (X)			
8.7	14.2	18.3	18.4
23.2	25.9	29.7	35.2
51.2	54.7	65.9	75

That means, the 25% one, the 50% one, and the 75% one.

DeepLearning.AI

Box-Plots

$$Q1 = q_{0.25} = \frac{18.3 + 18.4}{2} = 18.35$$

$$Q2 = q_{0.50} = \frac{25.9 + 29.7}{2} = 27.8$$

$$Q3 = q_{0.75} = \frac{51.2 + 54.7}{2} = 52.95$$

Interquartile range (IQR)

$$IQR = Q3 - Q1 = 52.95 - 18.35 = 34.6$$

Newspaper Advertisement (X)			
8.7	14.2	18.3	18.4
23.2	25.9	29.7	35.2
51.2	54.7	65.9	75

2

DeepLearning.AI

Box-Plots

$$Q1 = q_{0.25} = \frac{18.3 + 18.4}{2} = 18.35$$

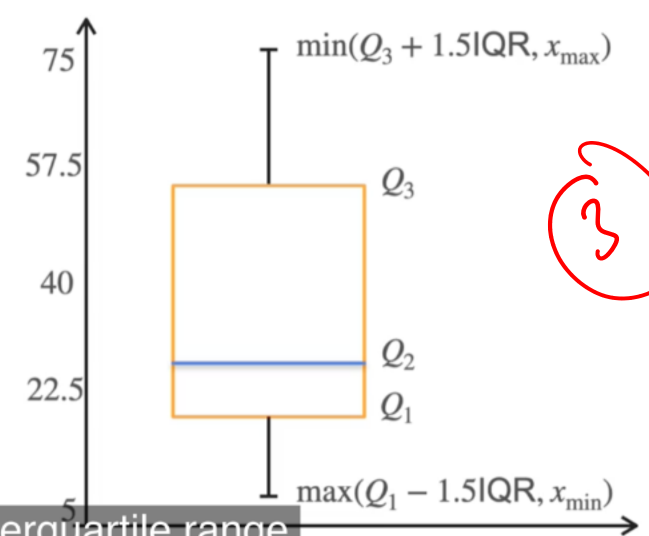
$$Q2 = q_{0.5} = \frac{25.9 + 29.7}{2} = 27.8$$

$$Q3 = q_{0.75} = \frac{51.2 + 54.7}{2} = 52.95$$

Interquartile range (IQR)

$$IQR = Q3 - Q1 = 52.95 - 18.35 = 34.6$$

$x_{min} = 8.7$ $x_{max} = 75$



3

1.5 times the interquartile range, which is Q3-Q1.

DeepLearning.AI

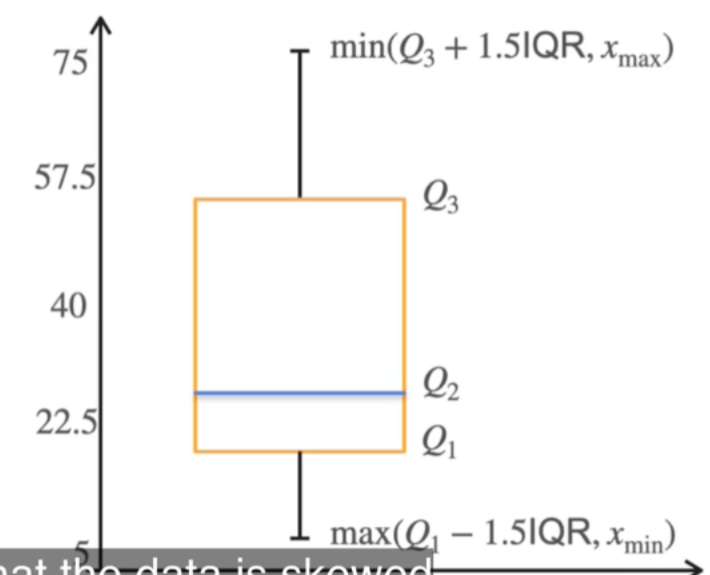
Box-plots gives analysis and insights of dispersions.

Box-Plots

4

What can you tell from this plot?

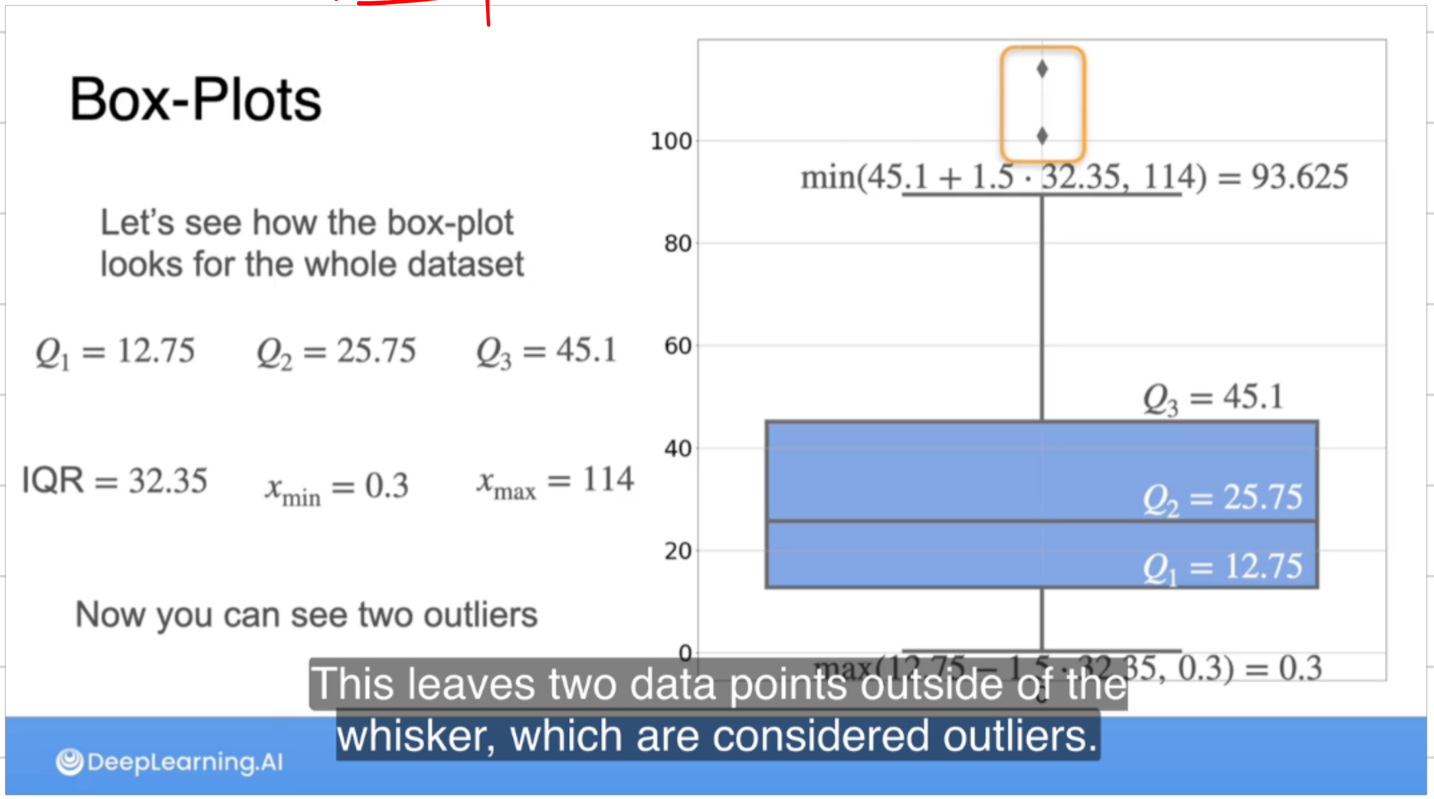
- Data is skewed
- No outliers (whiskers were cut at max and min value)
- Analyze dispersion

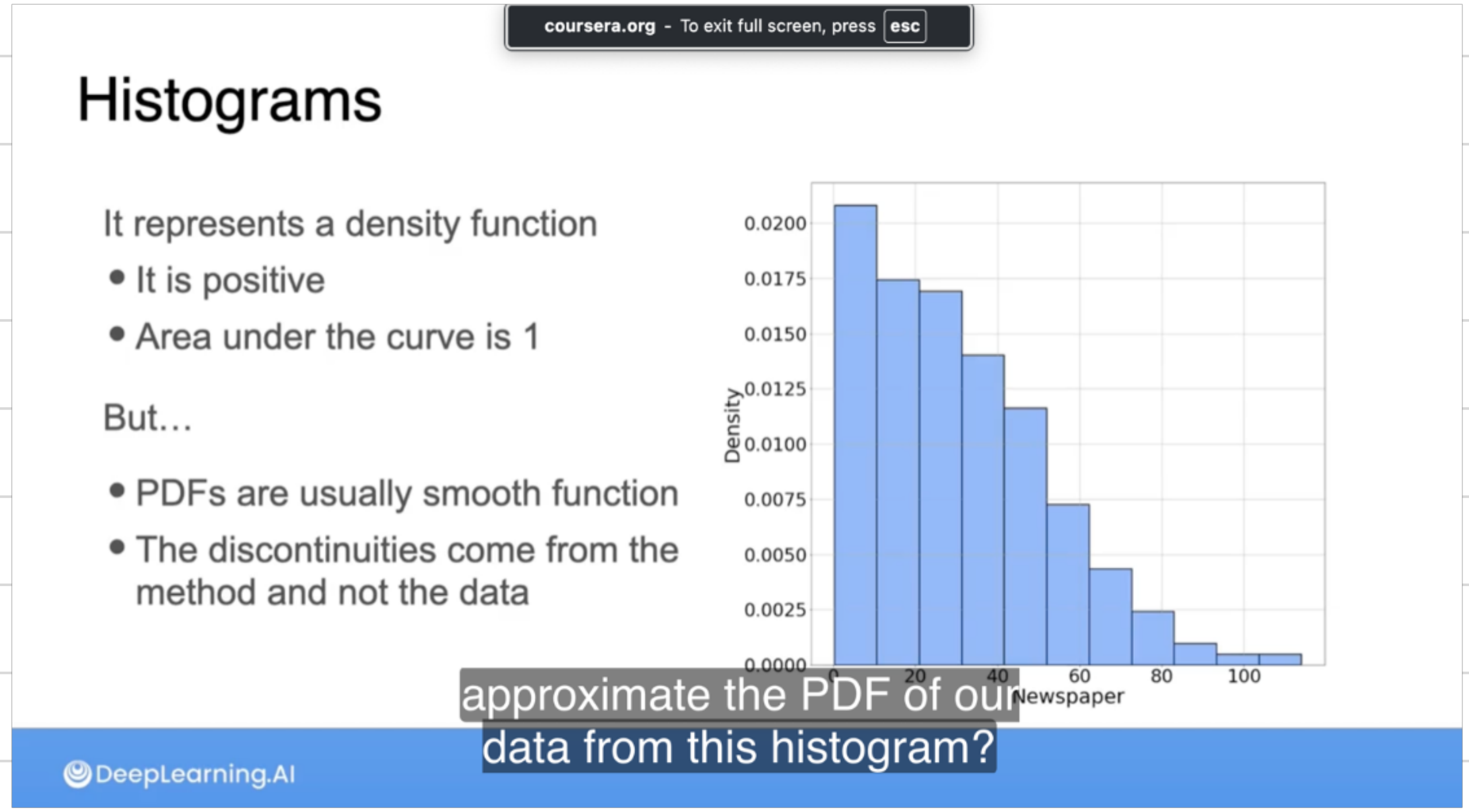


you can easily see that the data is skewed because Q3-Q2 is way bigger than Q2-Q1.

DeepLearning.AI

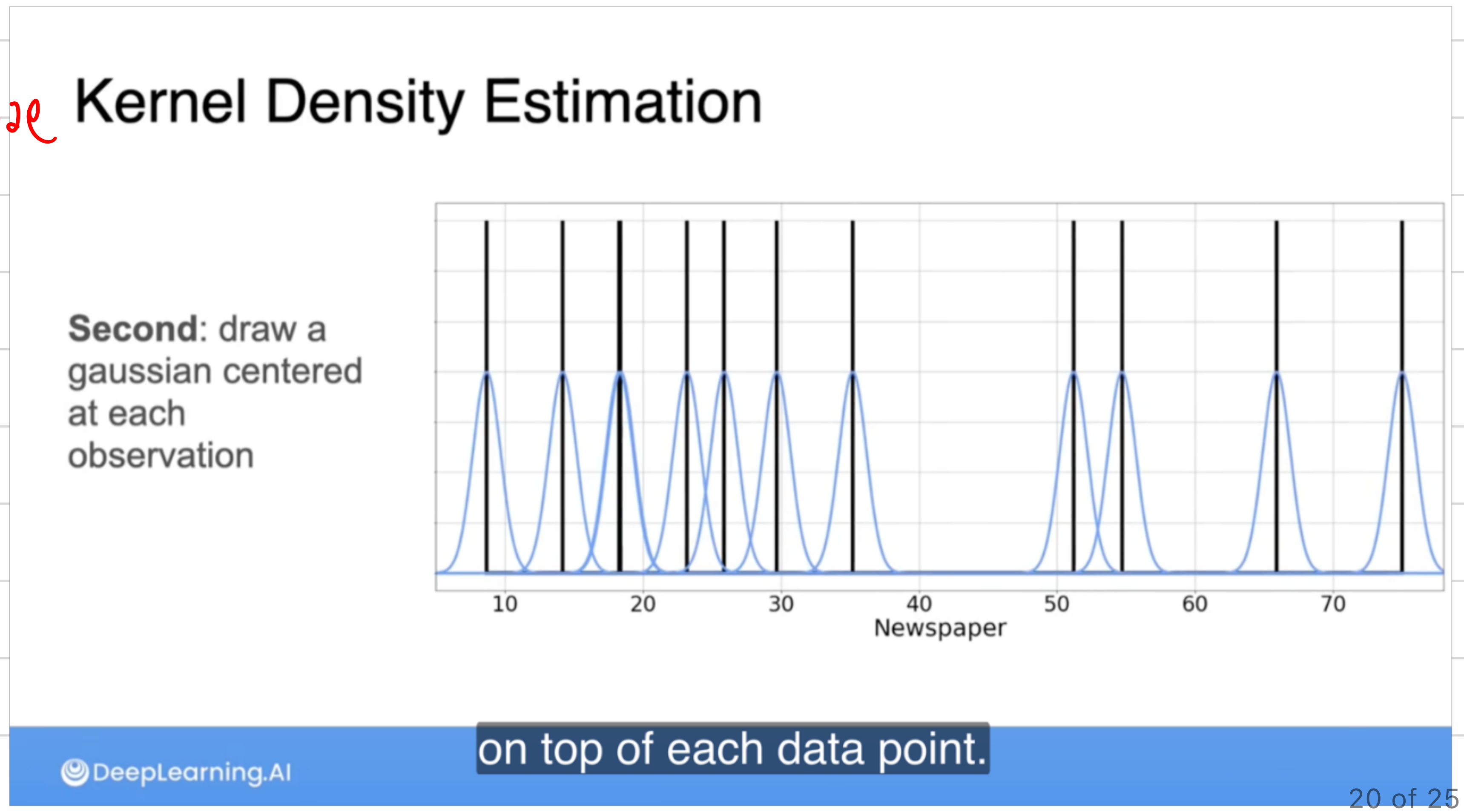
Box plots.





We can use KDE to visulize dataset without assumming any underlyings.

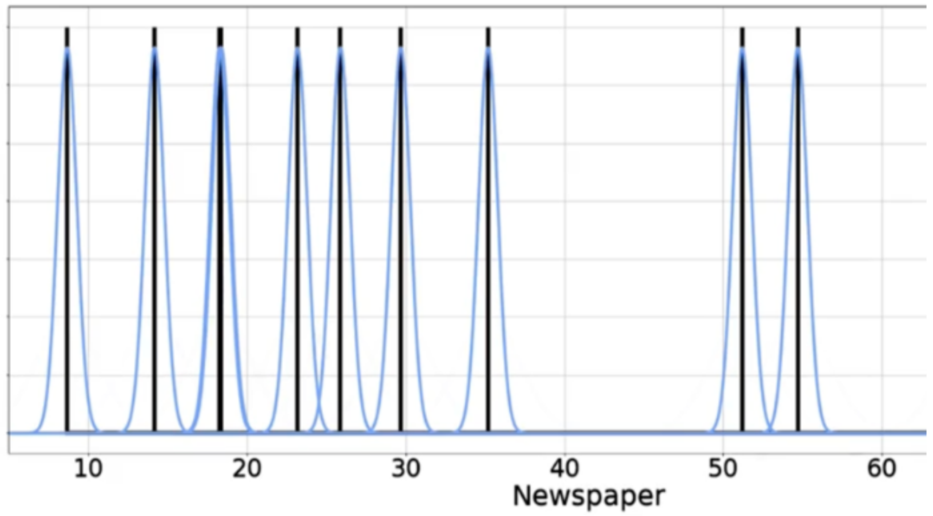
Non-parametric way to estimate PDF of a random-variable.



Kernel Density Estimation

1

Second: draw a gaussian centered at each observation



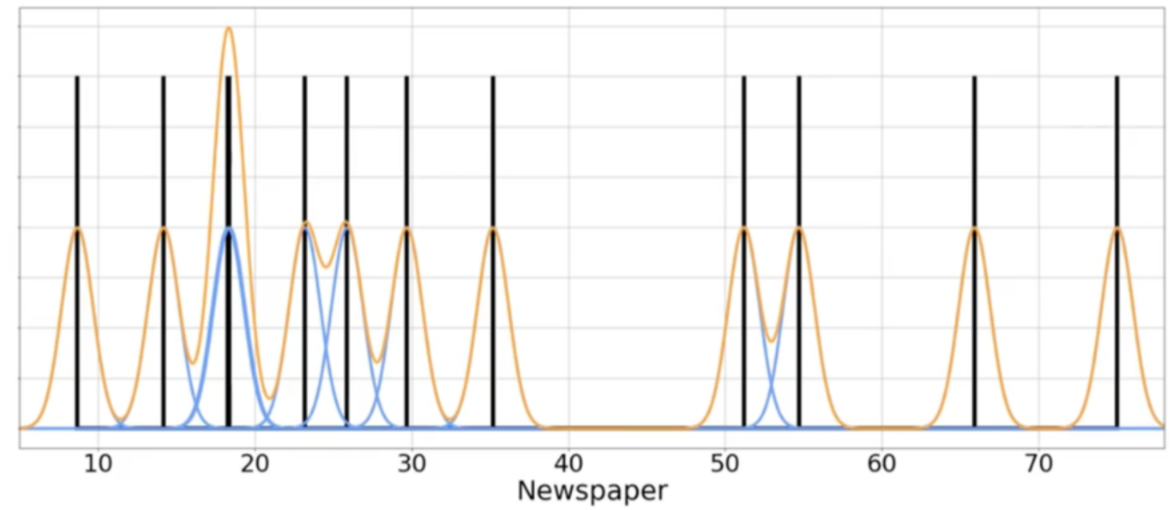
It could be thin or it could be wide.

DeepLearning.AI

Kernel Density Estimation

2

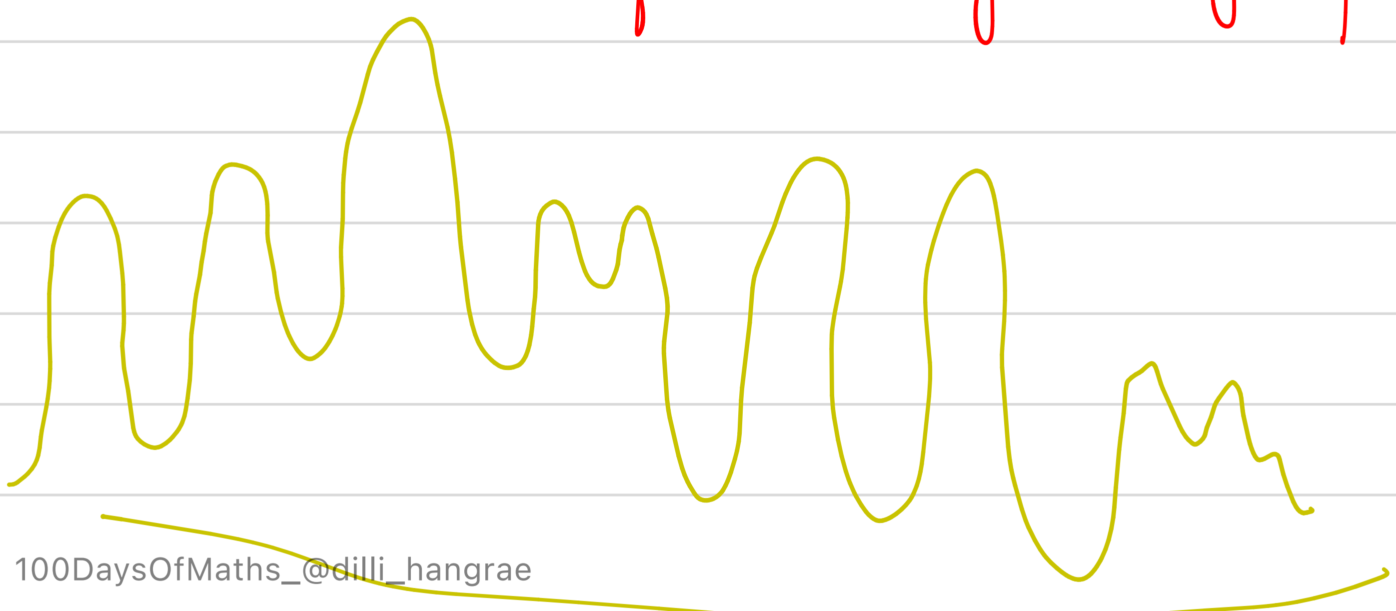
Third: multiply everything by 1/n and sum the curves



This is what estimation

DeepLearning.AI

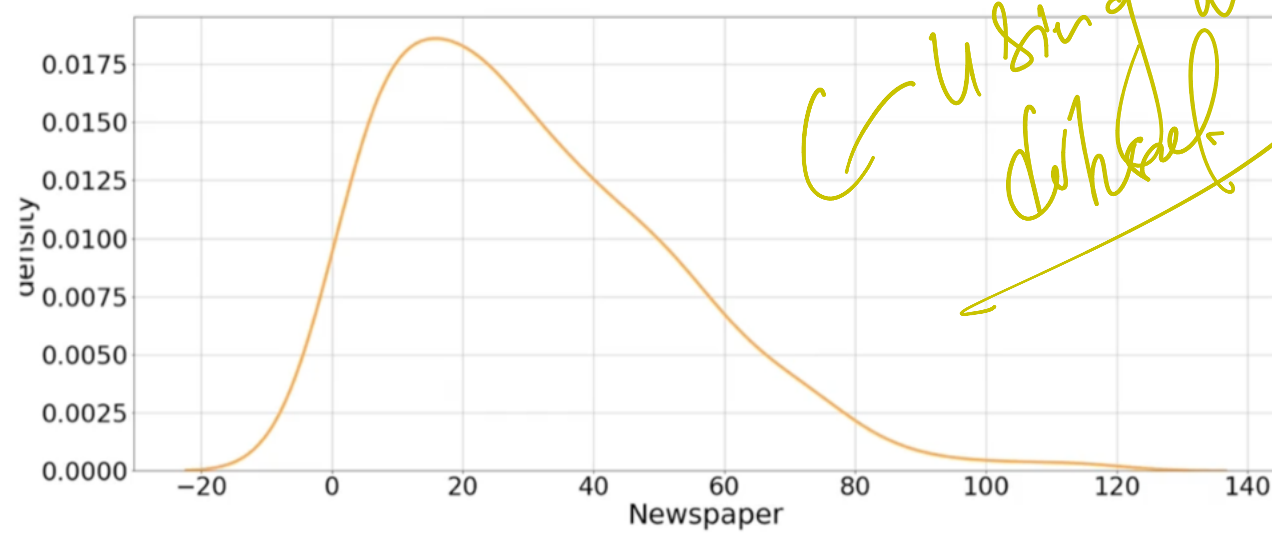
Now we if draw only Orange part.



Kernel Density Estimation

4

What if you used all the dataset?

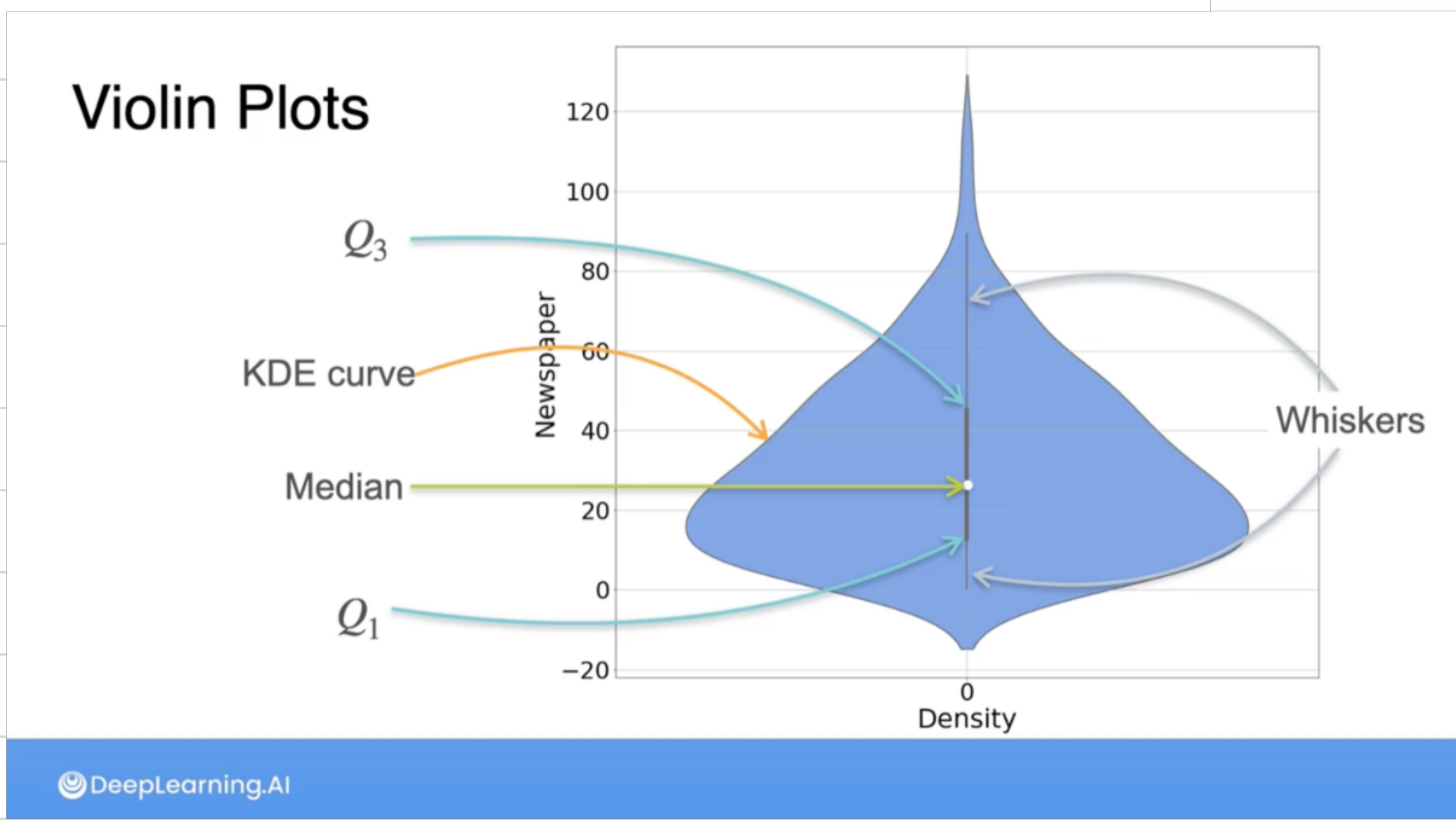


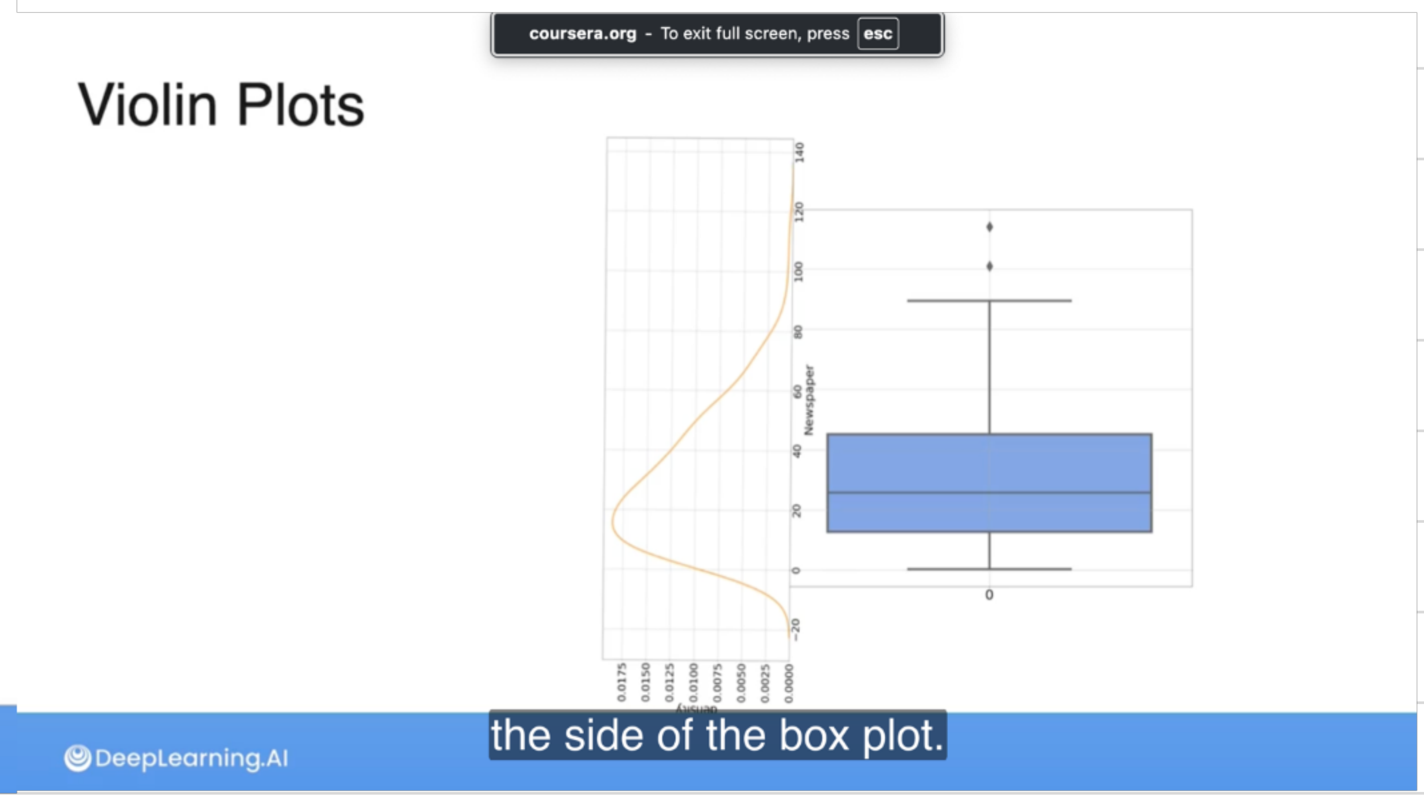
Using all dataset

This is a way to approximate the PDF based on the data.

DeepLearning.AI

Violin plots contains Information about box plots and KDE.





Models that Assumes the variable follows gaussian distribution.

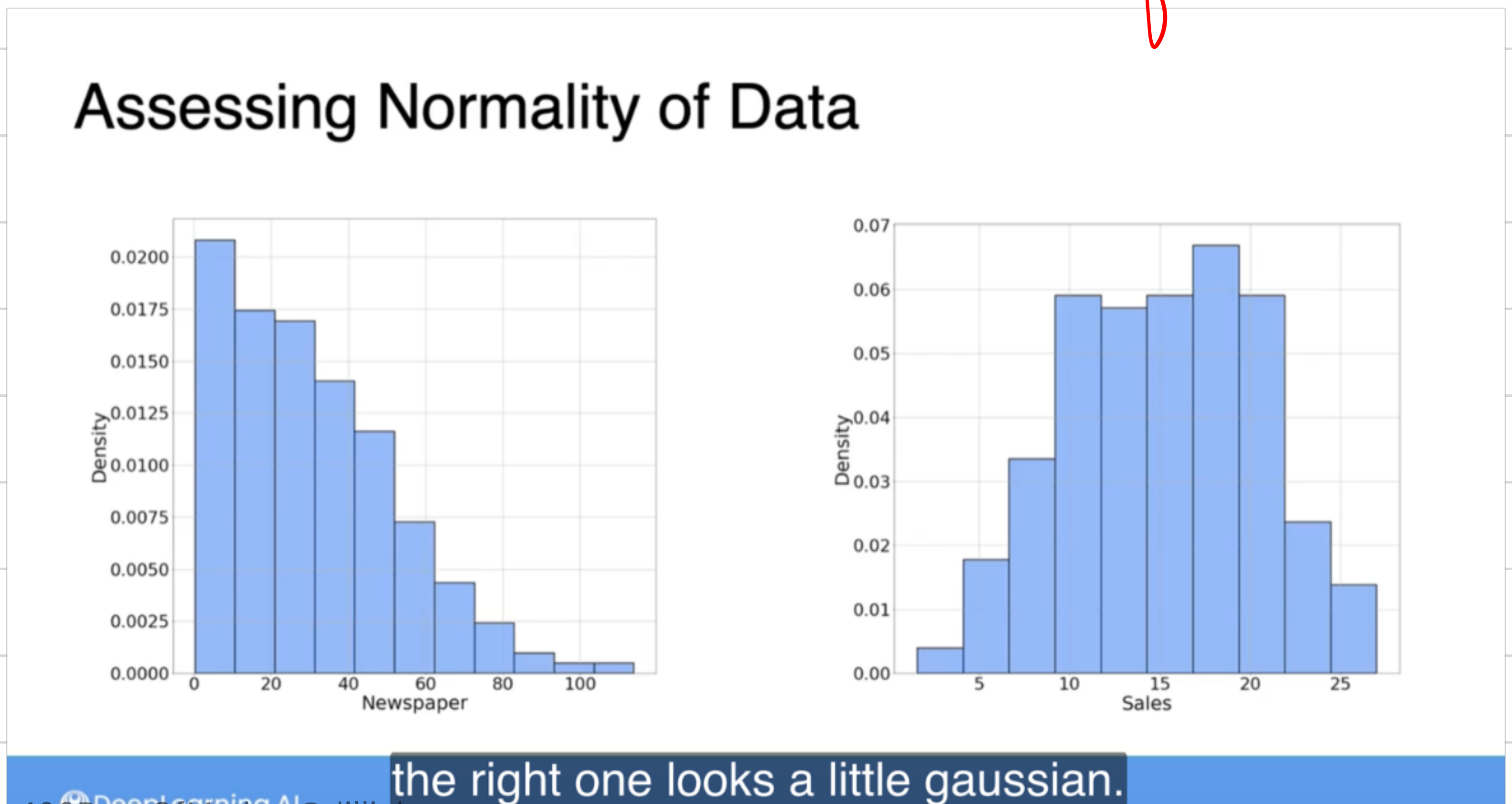
- Linear Regression
- Logistic Regression
- Gausian Naive Bayes
- Others

Some DS Assume Data Normality

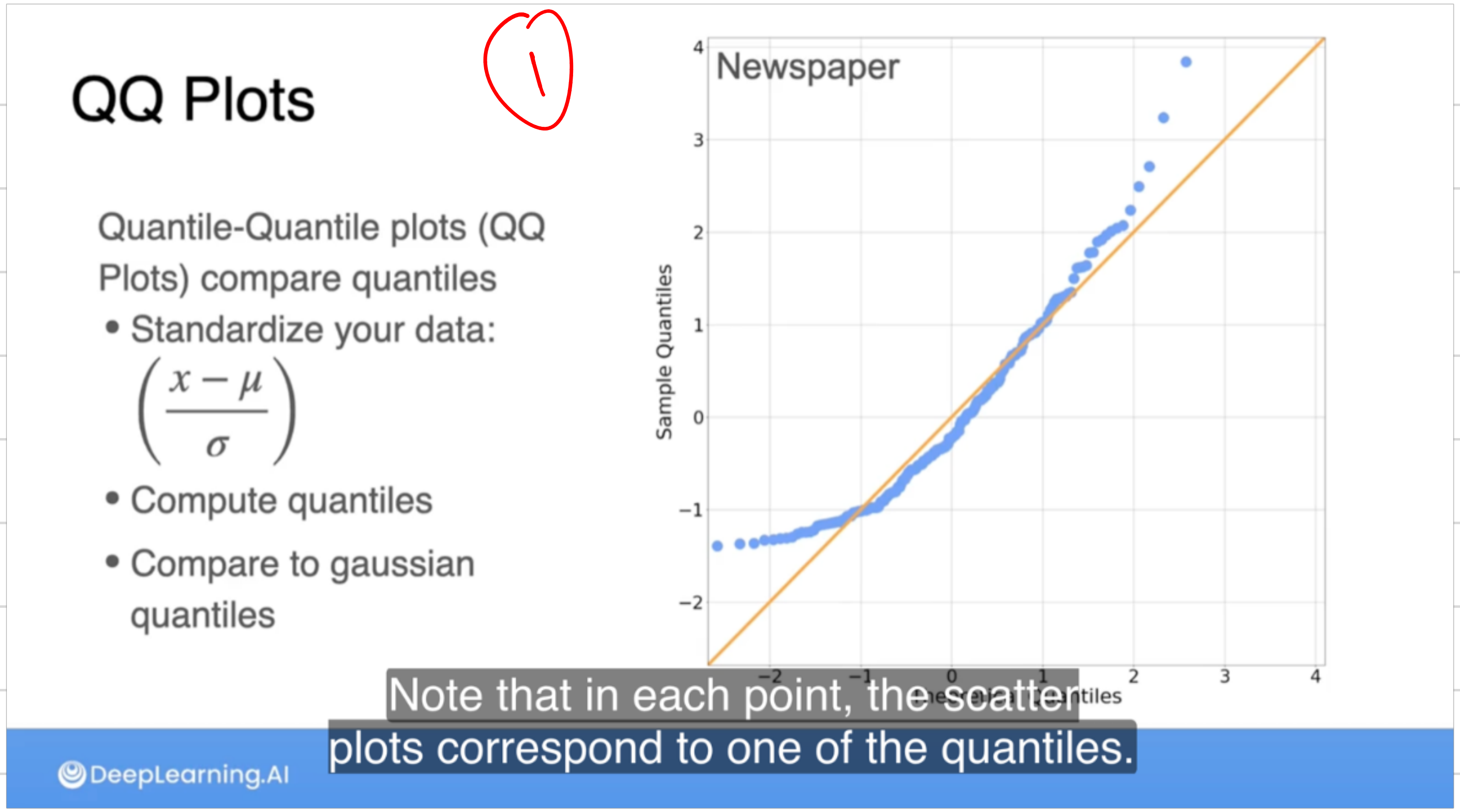
How do we check if the data distribution is Gaussian or Not?

Answer → Use QQ for that

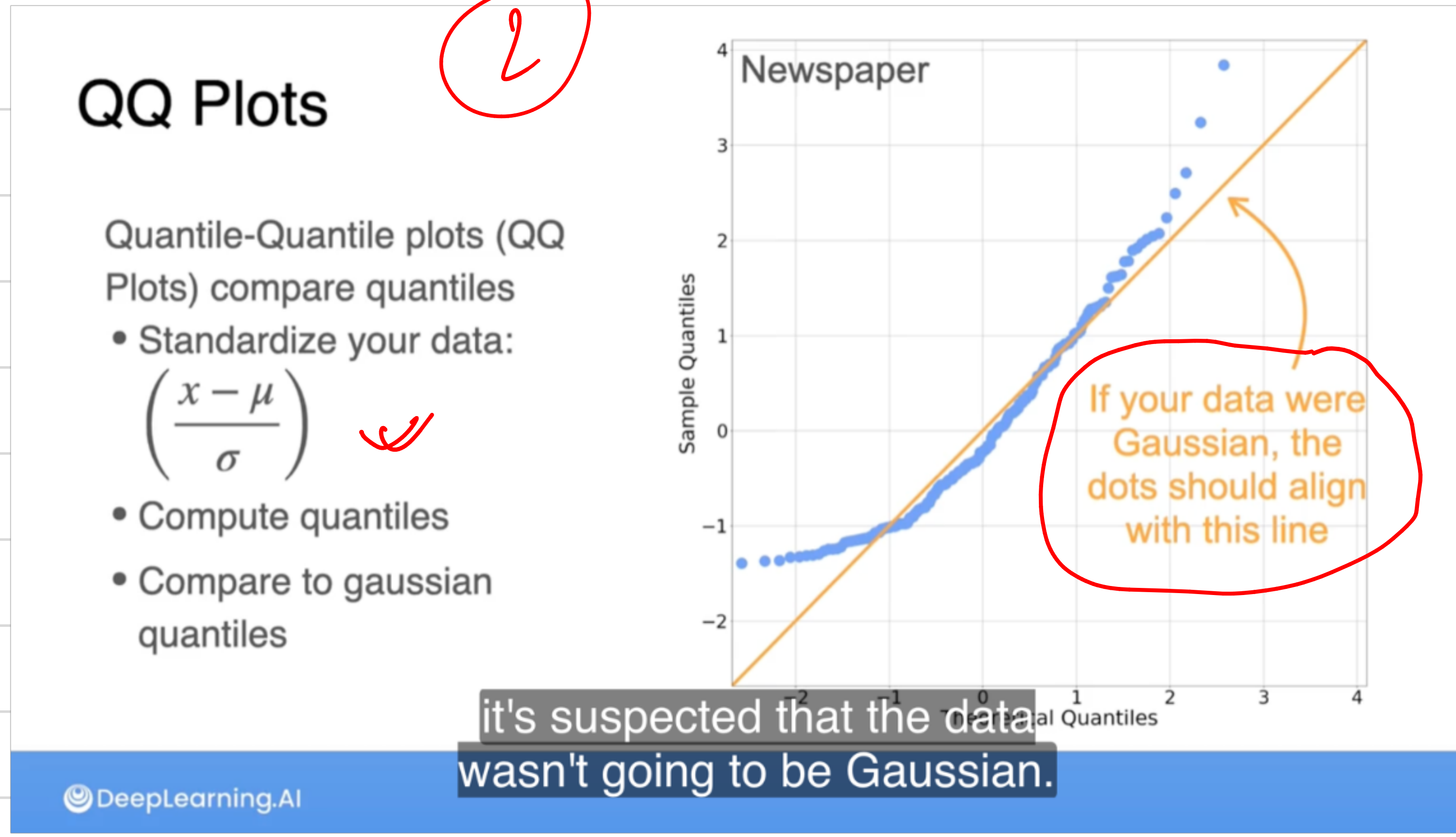
$\left(\frac{x-\mu}{\sigma}\right)$ Standardize the data.

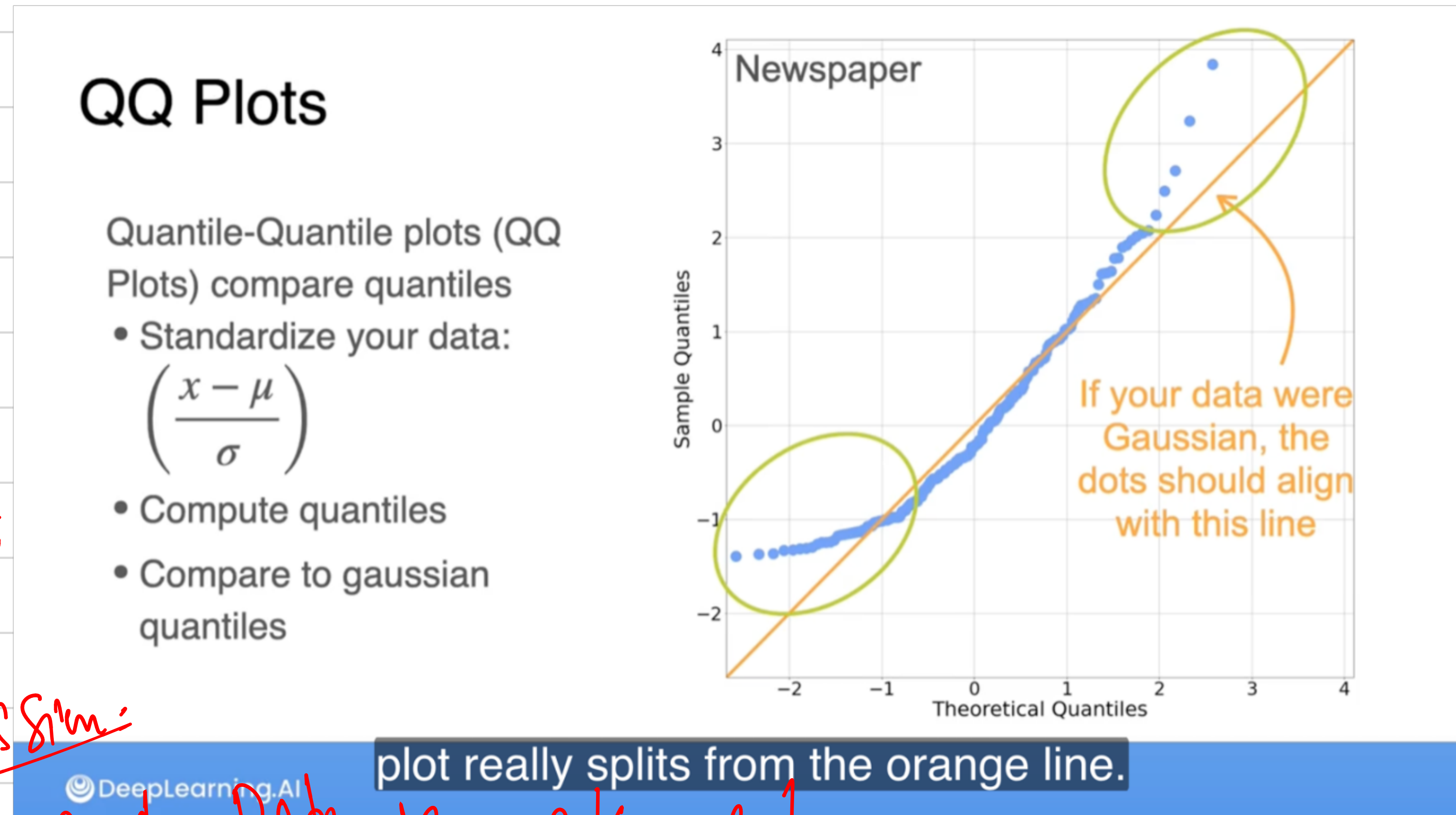


1



2

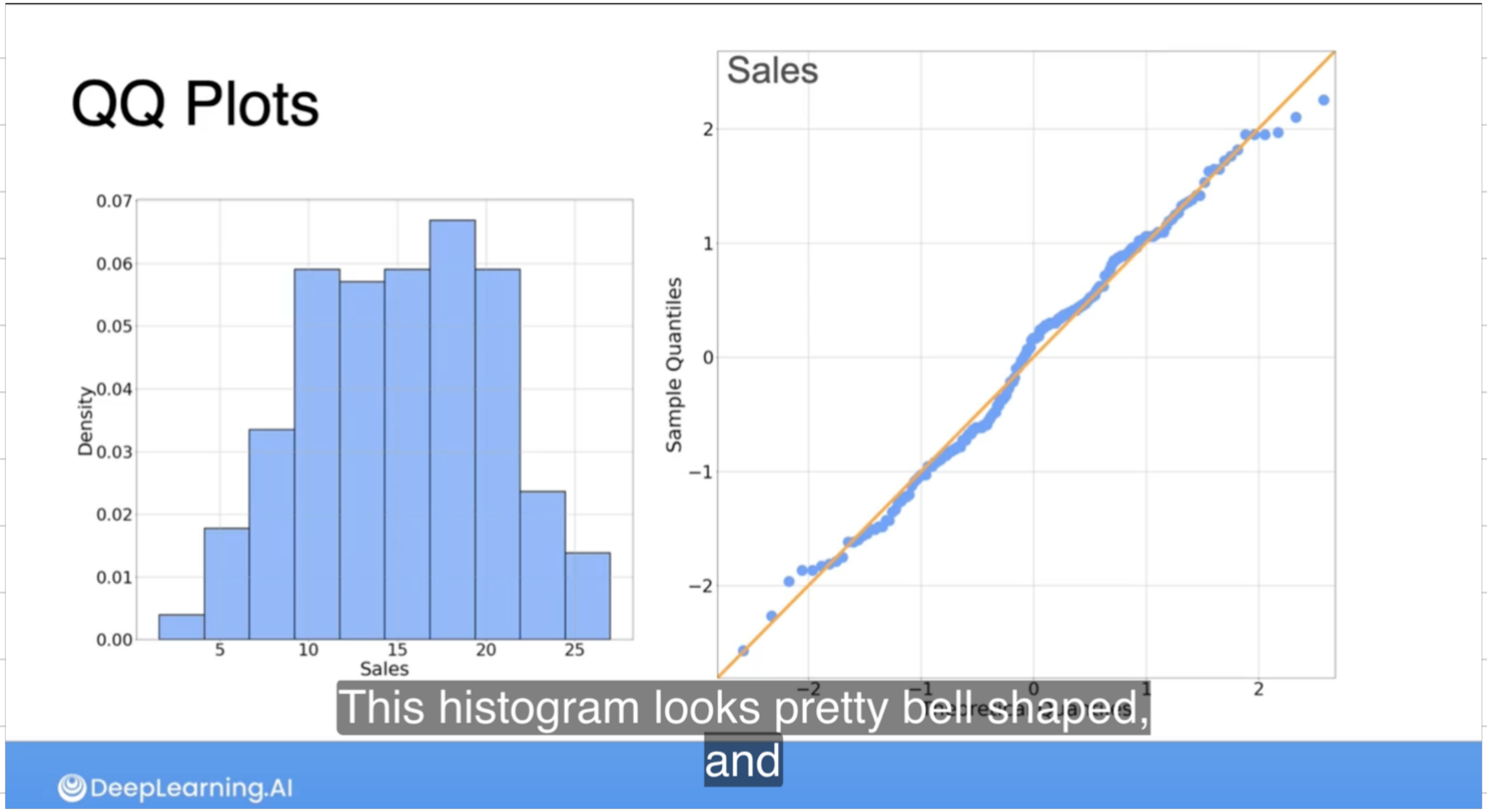




plot must have orange line with dots nearing it. The two circles shows data doesn't align straigh with line the plot suspected to be not Gaussian and Data is skewed.

This is the perfect example of Gaussian Distribution.

Since the line has surrounding dots



Source: Coursera Probability & statistics for Data Science & Machine Learning.