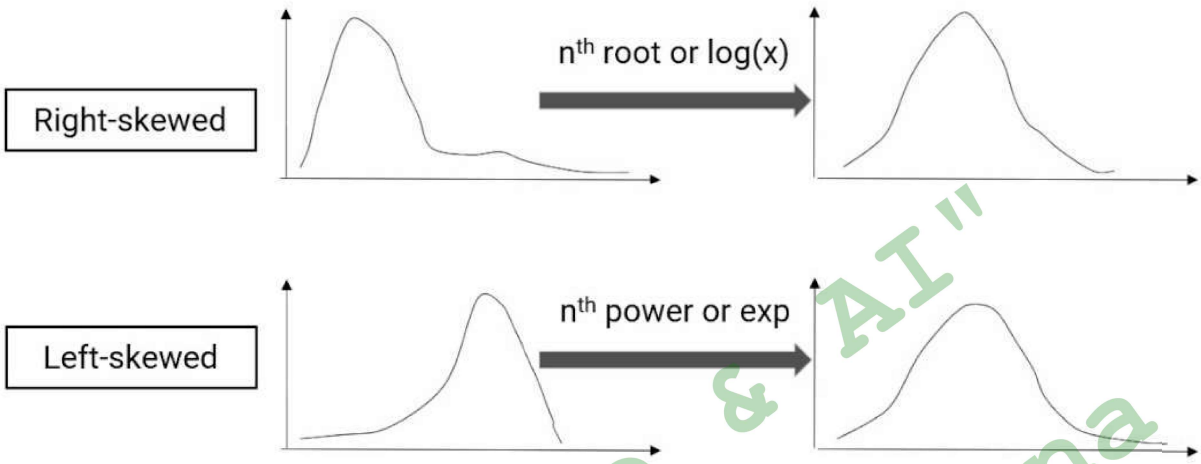
Feature Preprocessing: Transformatio



In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
data=pd.read_csv('titanic.csv')
data.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Na
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Na
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Na

In [3]:
```
1 data=pd.read_csv('titanic.csv',usecols=['Fare'])
2 data.head()
```

Out[3]:

	Fare
0	7.2500
1	71.2833
2	7.9250
3	53.1000
4	8.0500

In [4]:
```
1 data['Fare'].skew()
```

Out[4]: 4.787316519674893

Root Transformation

In [5]:
```
1 data['sqr_Fare']=data['Fare']**(1/4)
2 data['sqr_Fare'].skew()
```

Out[5]: 0.5196788882063811

Logarithmic Transformation

In [6]:
```
1 data['Log_Fare']=np.log(data['Fare']+0.01)
2 data['Log_Fare'].skew()
```

Out[6]: -2.4100496984507678

Reciprocal transformation

In [7]:
```
1 data['Rec_Fare']=1/(data['Fare']+0.01)
2 data['Rec_Fare'].skew()
```

Out[7]: 7.523650082079874

BoxCox

The Box-Cox transformation is defined as:

(X**λ − 1)/λ

where Y is the response variable and λ is the transformation parameter. λ varies from -
the transformation, all values of λ are considered and the optimal value for a given vari
selected.

In [8]:
```
1 from scipy import stats
```

In [9]:
```
1 # transform training data & save lambda value
2 data['Fare_boxcox'], param = stats.boxcox(data.Fare+0.01)
```

In [10]:
```
1 print("λ =",param)
```

```
λ = 0.18091321955494596
```

In [11]:
```
1 data['Fare_boxcox'].skew()
```

Out[11]:
```
0.25282461235568676
```