Let's see what is standardization and normalization,as it is a intergral part of feature scaling.

Suppose when we collect the data,we have a lot of features and those features includes your independent features and the dependent features.

With the help of the independent features we try to predict dependent feature and it is in supervised machine learning.

Now,basically these features have two important properties.

1)UNITS

2)MAGNITUDE

Suppose we have a data and it includes height,weight and so on and we want to find out Age(dependent feature)and in this Age has unit(no. of

years)and magnitude will be the numbers(values).Also,Data has many features and it differ from others,so this unit and magnitude need to be

scaled down to a particular scale.

TWO IMPORTANT SCALING TECHNIQUES USED ARE:

1)STANDARDIZATION

2)NORMALIZATION

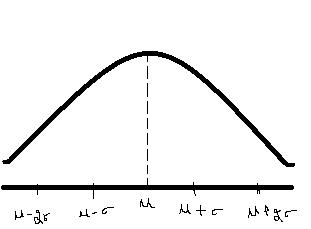
1)NORMALIZATION(also called min max scalar):It helps you to scale down your feature between '0' and '1'.

2)STANDARDIZATION:It helps you to scale down your feature based on standard normal distribution

Q)WHAT IS STANDARD NORMAL DISTRIBUTION?

ANS)Over here,it means that my mean is '0' and standard deviation is '1'.

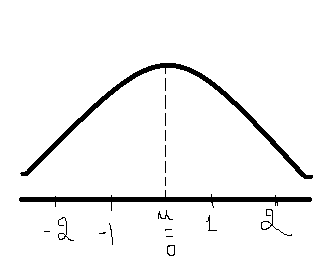
Normal distribution and the Gaussian distribution normally have bell like structure.



Over here centre element is mean(µ) and if I go right,it will be µ+σ(mean+S.D) and similarly If I go left it will be (µ-σ) and so on.

And if I want to convert it to standard normal distribution,here mean is ‘zero’, and σ =’1’

And if I go right it will be 2,similarly to left is -1 and -2 and so on.



Also we need to understand an imperical formula with respect to gaussian normal distribution is that within the 1st std.deviation ,there are around 68% of the total data whenever we have Gaussian normal distribution..

And for the 2nd standard deviation we have around 95% of the total distribution and for the third we have 99% of the data.

Now,we need to convert this whole into standard normal distribution,

We use a formula,

z-score = (Xi - µ)/ σ

where mean is ‘0’ and S.D is ‘1’.

Q)BUT WHY WE NEED TO CONVERT THIS?

ANS)An example,

Suppose,we have , {1,2,3,4,5} in our distribution and if we find the mean it will be µ = 3 .

Now,the formula,suppose we consider for 3rd value which is 3

Xi=3 , µ = 3, σ=1

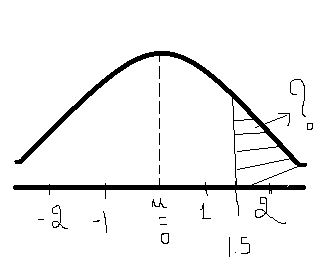
(3-3)/1 = 0, and similarly for the others.

Now,importsnt step,we know most of the values lie in 68% for 1 and so on.

Now,I want to find out that within 1.5 standard deviation away from the mean,I want to know the distribution.

Because of the imperical formula doesnot helps in finding.

Diagram:



For specifically this type of situation ,we use z-score . and also use z-score -table.

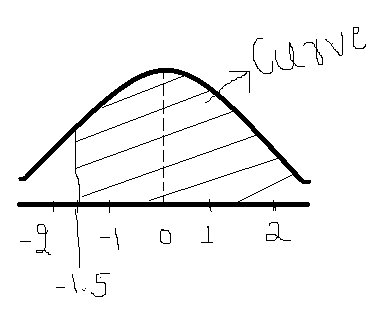
EXAMPLE: A population of class, where µ=75 and S.D =10

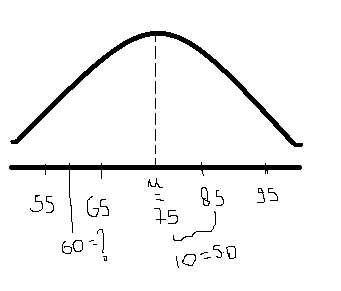
Q)To find probability where a student (x) score >60

P(x>60)

Diagram 1 and trying to convert it into Diagram 2

S.N.D ,here , 75 becomes 0,





65 as -1 ,85 =1 ,and 60 as -1.5.

Now the question, i.e; >60 so,I need to find this curve.

With empirical formula we cannot find out , so, we use z-score table.

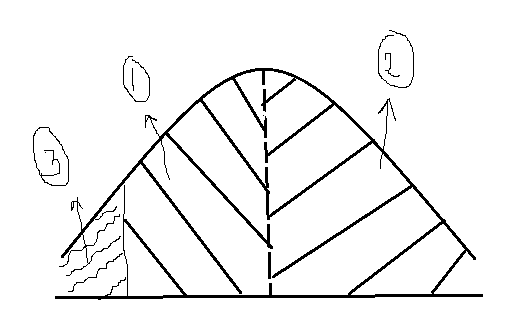
Q)what is this z-score table?

ANS) It basically says that what will be the distribution below -1.5 .

Now, in order to compute this, we will divide this curve, (entire area) into 3 regions.

1)Region 1 2)Region 2 3)Region 3

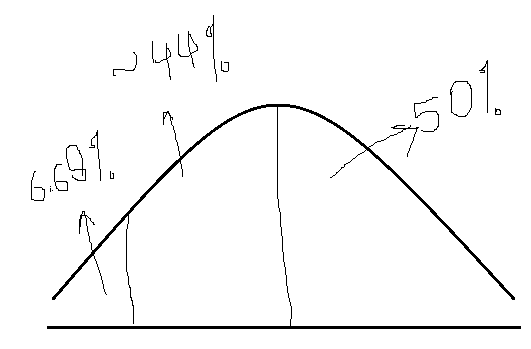
Diagram:



z-score of -1.5 is 0.0668 i.e ; region 3 is 0.0668,

Now for region 2, Diagram :

In S.N.D ,we know ratio is divided into symmetrical ,



Which means region 2 is 50%

Whole area = x+50%+6.68

X=100-56.68

X=44% approx. for region 1,

Now ,the question ,I will just combine

44+50 =94%

So,probability is 94%