Logistic Regression - Special type of Generalized linear model because it predict categories. LR fits "S" shaped logistic function. Curves goes from 0 to 1. For eg, Customer will invest in FD? (yes/No). So from 0 to 1, we can infer what is the probability customer will invest in FD.

Advantage of LR - No assumptions about distribution of target class, model interpretation is easy, less inclined to overfitting but can overfit in high dimensional data. L1 and L2 method are there to avoid overfit.

Disadvantages of LR - High dimensional data, LR tends to overfit, Sensitive to Outliers, No ~~corr~~ correlation which lead to multicollinearity.

Probability vs Likelihood - Probability is the percentage that a success occur, Eg, tossing a coin and winning the head lead to 0.5 probability.

Likelihood is a conditional probability of a event (set of success) occur by knowing the probability of a success occur. Eg, toss coined 10 times and suppose we got 7 success and 3 failed. So, Likelihood $(0.5 | 7) = 0.1171$.

It suggest 0.1171 is the probability that above event will happen (7 success out of 10 trials) by knowing probability of one success is 0.5

Another eg, in cricket match, coin is tossed and one captain calls head and win the toss.

Now what is the probability winning captain will elect to bat? Probability $= \frac{1}{2}$ (bat or bowl)

Probability is straight 50% because he/she have 2 options either bat or bowl.

But if question is likelihood that winning captain will elect to bat, then it will not be straight 50% because likelihood depend on type of pitch, strength, weakness etc.

So probability is straight up number but likelihood is a function of many parameters, conditions

Probability vs Odd - $\text{Probability} = \frac{\text{something happening}}{\text{everything that could happen}}$ $\quad \text{Odds} = \frac{\text{Something happening}}{\text{Something not happening}}$.

Suppose we toss coin, outcome was 5 win, 3 loss. $\text{Odds(win)} = \frac{5}{3}$ $\quad \text{Probabity (win)} = \frac{5}{8}$

So, $\text{odds (win)} = \frac{\text{Probability (win)}}{\text{Probability (loss)}} = \frac{P(\text{win})}{1 - P(\text{win})}$, In short $\text{odds} = \frac{P}{1-P}$

Log(Odds) - Log of odds solve the problem of symmetry. Log(Odds) also known as Logit fn.

Suppose team is bad. $\text{Odd(win)} = \frac{P(\text{win})}{P(\text{loss})}$, Suppose Odd $= \frac{1}{4} = 0.25$, more worse $= \frac{1}{8} = 0.125$

more worse $= \frac{1}{16} = 0.0.06$, full worst $= \frac{0}{\text{anything}} = 0$

Suppose team is good, $\text{Odd(win)} = \frac{5}{3} = 1.7$, improve $= \frac{9}{3} = 3$, more improve $= \frac{27}{3} = 9$, best $= \infty$.

Odd range (loss) will be 0 to 1 (Denominator > Numerator)

Odd range (win) will be 1 to ∞ (Denominator < Numerator)

So log(odds) make the data symmetric.

0 1 2 3 4 5 ∞

Odd(loose) Odd(win)

For eg, if odds are against 1 to 6, odds $= \frac{1}{6} = 0.17$, but in favour odds $= \frac{6}{1} = 6$.

Assymetric it is but if add log to it $\log(\frac{1}{6}) = -1.79$ and $\log(\frac{1}{6}) = 1.79$.

So things become symmetric -1.79 and 1.79

$\log(\text{odds}) = \log\left(\frac{P}{1-P}\right)$, this is also known as Logit function which is log of ratio of probabilities and log make range $-\infty$ to ∞ and midpoint is 0

Odds Ratio and log (Odd Ratio) = Odd Ratio = $\frac{Odd_1}{Odd_2} = \frac{\overset{xx}{0000}}{000} = \frac{2/4}{3/1} = 0.17$

If denominator > numerator, odd ratio goes from 0 to 1.

And if denominator < numerator, odd ratio goes from 1 to ∞. Taking log make symmetrical.

Odd Ratio tells us if there is strong / weak relationship between two variables.

Eg →

		Invested FD Yes	No
Age	High	60	100
	Low	20	100

Given a person' high age, Odd (FD) = 60/100

Given a person low age, Odd (FD) = 20/100

$$\text{Odd (FD)} = \frac{60/100}{20/100} = \frac{0.6}{0.2} = 3$$

So odds ratio is suggest, 3 times greater that someone with higher age will invest in FD.

Larger the value of Odds, it is good predictor. Smaller the value of odds, it is not good predictor.

Another advantage, if we take random numbers, suppose all those add to 100 and calculate Odds of each number, it will not be normal distributed. so if we log(odds) it will become normally distributed.

~~Logit~~ Logistic Response function → Normally, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$.

In logistic we get probability, so $p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$ (i)

But p does not ensure it will be in 0 to 1, which is must for any probability because probability ranges from 0 to 1.

So apply sigmoid f^n and ensure p stays from 0 to 1. So $p = \frac{1}{1+e^{-p'}}$

Suppose, p' value goes to +infinity then P value will become 1 and if p' values goes to -infinity then P value become 0.

In this way, p is ranged from 0 to 1. Normally sigmoid $f^n = \frac{1}{1+e^{-z}}$.

$$\frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\dots+\beta_n x_n)}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad \text{— (i)}$$

To get exponential out of denominator, we use odds instead of probability.

$\text{Odds} = \frac{P}{1-P}$, to obtain probability from Odds, $P = \frac{\text{Odds}}{1+\text{Odds}}$ — (ii)

From eqn (i) and (ii) ⇒ $\text{Odds}(Y=1) = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\dots+\beta_n x_n}$

taking log on both side ⇒ $\log(\text{Odds}(y=1)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$

So log(odds) f^n, also known as logit f^n maps probability from (0, 1) to any value $(-\infty, \infty)$