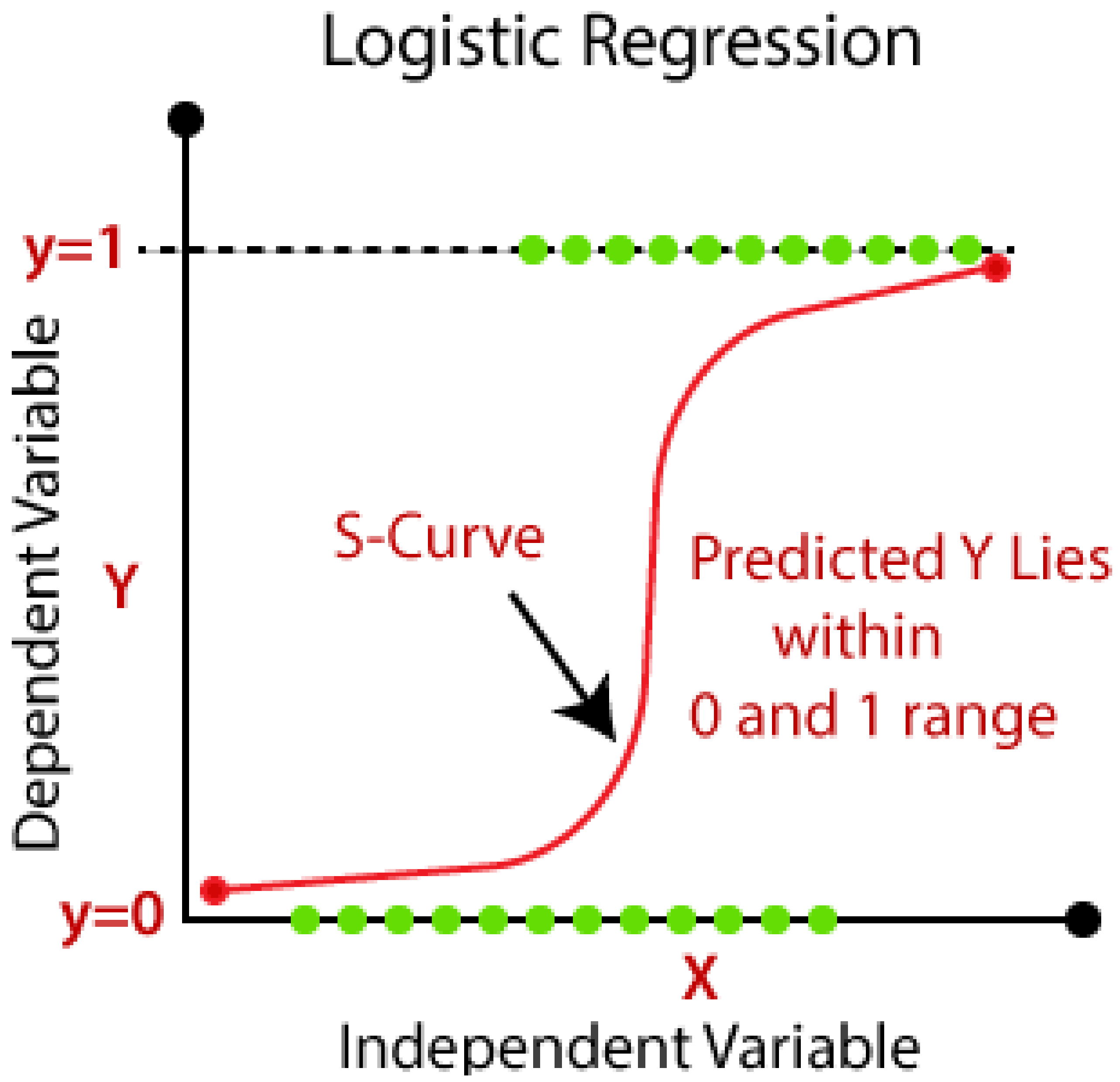
MV Ramarao

LOGISTIC REGRESSION THEORY

MV Ramarao

Topics Outline

1. Logistic Regression Introduction
2. Why we don't use Linear Regression ?
3. How Logistic Regression Works ?
4. Why we don't use MSE ?
5. Basic Assumptions of Logistic Regression
6. Advantages and Disadvantages

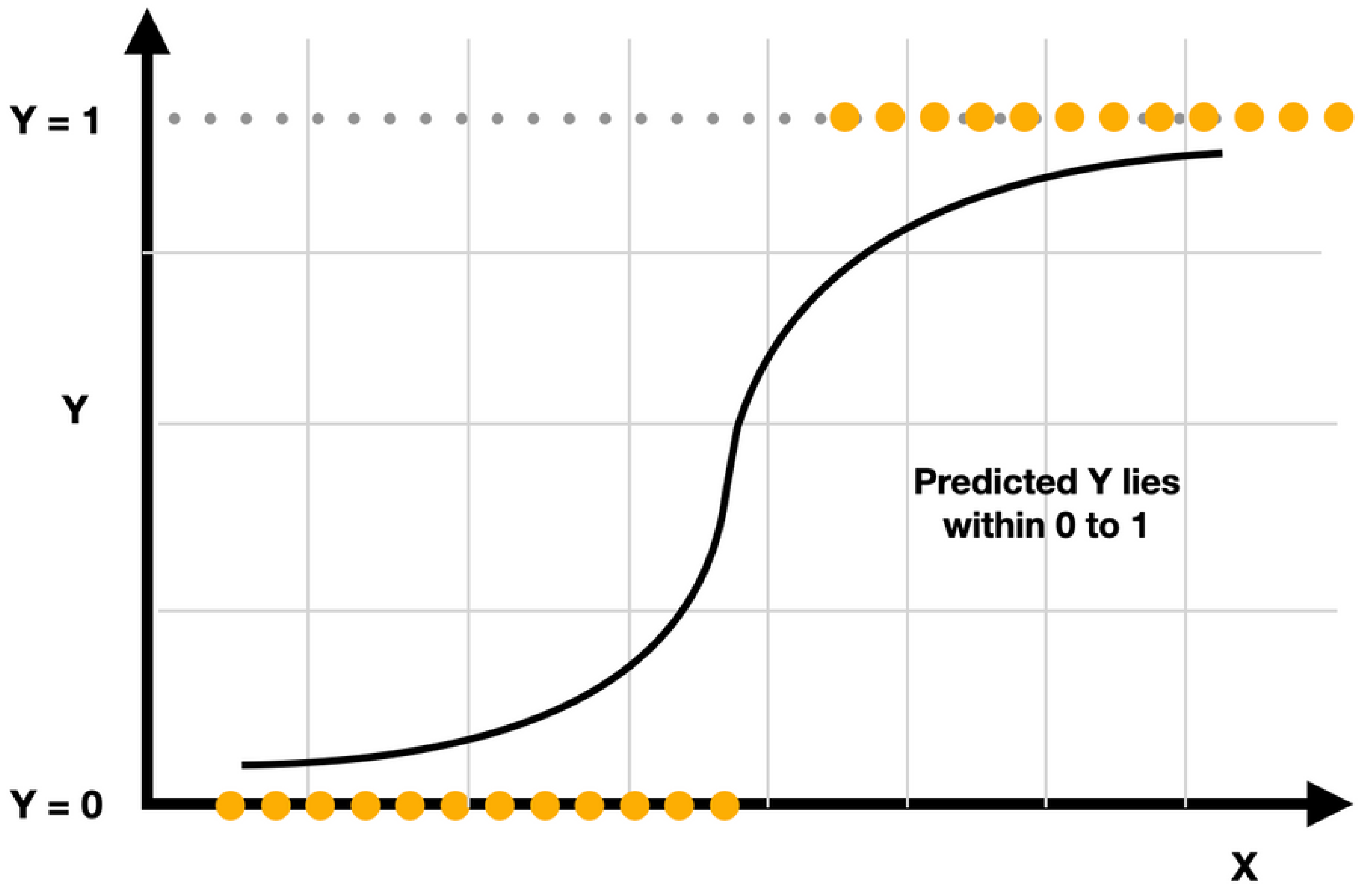


Logistic Regression Introduction

MV Ramarao

- Logistic Regression is one of the most common ML algorithms in supervised learning. It is used for predicting the categorical dependent variable using a given set of independent variables.

- It is a probabilistic classifier. It predicts the probability of a particular label. Using Threshold value and output probability we decide the label.

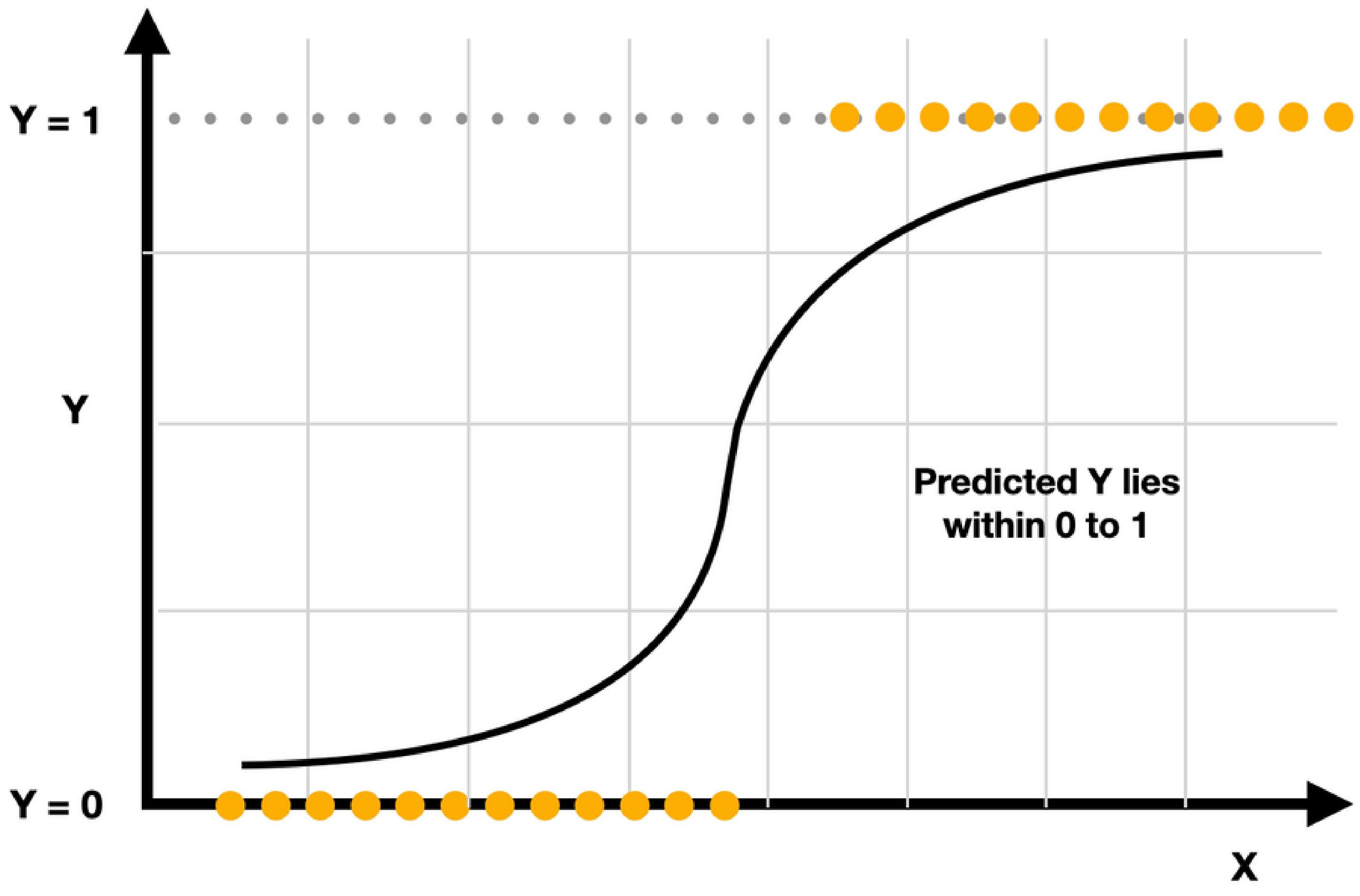


Logistic Regression Introduction

MV Ramarao

- Logistic Regression is much similar to the Linear Regression except that how they are used. **Linear Regression** is used for solving **Regression problems**, whereas **Logistic regression** is used for solving the **classification problems**.

- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).



Why we dont Use Linear Regression ?

MV Ramarao

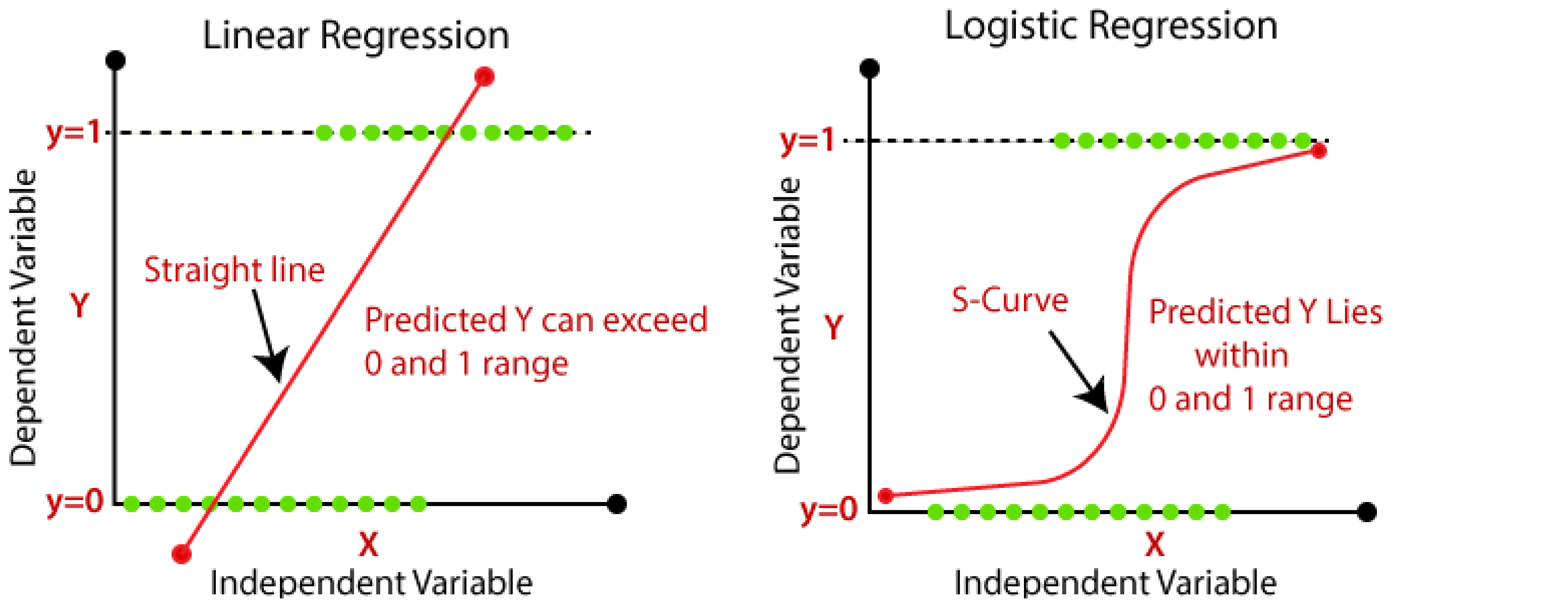
Two Main Reasons

- Linear regression is suitable for predicting output that is continuous value which ranges from -infinity to + infinity. In Classification tasks we need probabilities which lies within 0 to 1 to identify the category.
- Although we can convert the values which are <0 to 0 and values >1 to 1, it is not preferable.
- The second problem is regarding the shift in threshold value when new data points are added. This shows us the model is highly in stable and we cant afford these fluctuations in deployment.

How Logistic Regression Works ?

MV Ramarao

It works similar to Linear Regression. After finding the curve, we additionally apply logistic function (sigmoid function) on top of it to get probabilities.



How Logistic Regression Works ?

MV Ramarao

chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0.076	11.0	34.0	0.9978	3.51	0.56	9.4	bad
0.098	25.0	67.0	0.9968	3.2	0.68	9.8	bad
0.092	15.0	54.0	0.997	3.26	0.65	9.8	bad
0.075	17.0	60.0	0.998	3.16	0.58	9.8	good
0.076	11.0	34.0	0.9978	3.51	0.56	9.4	bad
0.075	13.0	40.0	0.9978	3.51	0.56	9.4	bad
0.069	15.0	59.0	0.9964	3.3	0.46	9.4	bad
0.065	15.0	21.0	0.9946	3.39	0.47	10.0	good
0.073	9.0	18.0	0.9968	3.36	0.57	9.5	good
0.071	17.0	102.0	0.9978	3.35	0.8	10.5	bad
0.097	15.0	65.0	0.9959	3.28	0.54	9.2	bad
0.071	17.0	102.0	0.9978	3.35	0.8	10.5	bad
0.08900000000000000	16.0	59.0	0.9943	3.58	0.52	9.9	bad
0.114	9.0	29.0	0.9974	3.26	1.56	9.1	bad
0.17600000000000000	52.0	145.0	0.9986	3.16	0.88	9.2	bad

Independent Variables

Dependent Variable

Step 1: Create Linear Function (Linear Regression)

Z = w1 * x1 + w2 * x2 + w3 * x3 + w4 * x4 + w5 * x5 + w6 * x6 + w7 * x7 + b

Step 2: Apply Sigmoid function on top of it

$$S(x) = \frac{1}{1 + e^{-z}}$$

How Logistic Regression Works ?

MV Ramarao

chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0.076	11.0	34.0	0.9978	3.51	0.56	9.4	bad
0.098	25.0	67.0	0.9968	3.2	0.68	9.8	bad
0.092	15.0	54.0	0.997	3.26	0.65	9.8	bad
0.075	17.0	60.0	0.998	3.16	0.58	9.8	good
0.076	11.0	34.0	0.9978	3.51	0.56	9.4	bad
0.075	13.0	40.0	0.9978	3.51	0.56	9.4	bad
0.069	15.0	59.0	0.9964	3.3	0.46	9.4	bad
0.065	15.0	21.0	0.9946	3.39	0.47	10.0	good
0.073	9.0	18.0	0.9968	3.36	0.57	9.5	good
0.071	17.0	102.0	0.9978	3.35	0.8	10.5	bad
0.097	15.0	65.0	0.9959	3.28	0.54	9.2	bad
0.071	17.0	102.0	0.9978	3.35	0.8	10.5	bad
0.08900000000000000	16.0	59.0	0.9943	3.58	0.52	9.9	bad
0.114	9.0	29.0	0.9974	3.26	1.56	9.1	bad
0.17600000000000000	52.0	145.0	0.9986	3.16	0.88	9.2	bad

Independent Variables

Dependent Variable

Step 3: Calculate the loss function using log loss

$$\text{Loss} = -y * \log(p) - (1-y) * \log(1-p)$$

Step 4: Use Gradient descent to Update the weights or coefficents and bias

$$w_{new} = w - \alpha * \frac{\delta L}{\delta w}$$

$$b_{new} = b - \alpha * \frac{\delta L}{\delta b}$$

Why we dont Use Mean Square Error (MSE) ?

- In Logistic regression, we often use gradient descent to find the optimal values for coefficients. If the loss function is not convex, it is not guaranteed that we will always reach the global minima. Majorly we will struck at local minima.

- The values we get while calculating MSE requires high storage as it results the values with high precision floating point numbers and these values are hard to interpret

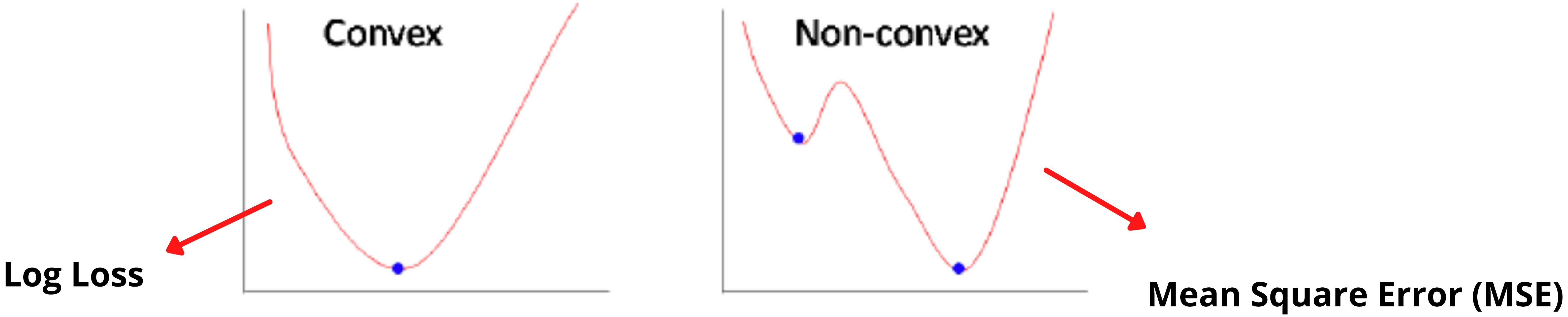


Figure 4: Convex and non-Convex functions

Basic Assumptions of Logistic Regression

MV Ramarao

- **Independence:** Logistic regression assumes that the observations in the dataset are independent of each other. That is, the observations should not come from repeated measurements of the same individual or be related to each other in any way.

- **Collinearity:** Logistic regression assumes that there is no severe multicollinearity among the explanatory variables

- **Linearity:** Logistic regression assumes that there exists a linear relationship between each explanatory variable and the logit of the response variable. Recall that the logit (log (p / 1 - p))

- Logistic regression assumes that the sample size of the dataset if large enough to draw valid conclusions from the fitted logistic regression model.

Advantages and Disadvantages

MV Ramarao

Advantages	Dis Advantages
• Easy to interpret and implement • Fast predictions for unknown records • Makes no assumptions about distributions of classes in feature space • It can easily extend to multiple classes and a natural probabilistic view of class predictions.	• Not good for nonlinear problems • It will not able to identify complex relationships in the data • Features should be average or less collinear • Not work well for large dimensional data

Additional Resources

- Logistic Regression (By Andrew Ng) - Link
- Logistic Regression - Javatpoint
- Logistic Regression Detailed Overview - Towards Data Science
- Why Linear Regression is not suitable for classification? - medium
- Hands-On Machine Learning with Scikit-Learn, Keras and Tensor Flow (Aurelien Geron) - link

THANK YOU