# Basic Terminologies

* **Label:** The thing that we are trying to predict. It is the variable ‘y’ in the case of linear regression.
* **Features:** They describe our data. It is the variable ‘x’ in the case of linear regression.
* **Sample:** An instance of data (Labeled and Unlabeled)
* **Hypothesis:** Function that maps the input features to the output label
* **Training:** Enabling the model to learn the relationship between features and label.
* **Testing:** Evaluating the function against unforeseen data
* **Parameters:** Coefficients in a hypothesis function
* **Hyperparameters:** Knobs that can be tweaked to increase the efficiency of the model
* **Vector:** One dimensional arrangement of elements
* **Matrix:** Two dimensional arrangement of elements
* **Tensor:** N-dimensional arrangement of elements.
* **Regression:** Continuous data
* **Classification:** Discrete data

# 

# Framework Magic

## Creating a model

model = tf.keras.Sequential([keras.layers.Dense(units=1, input\_shape=[1])])

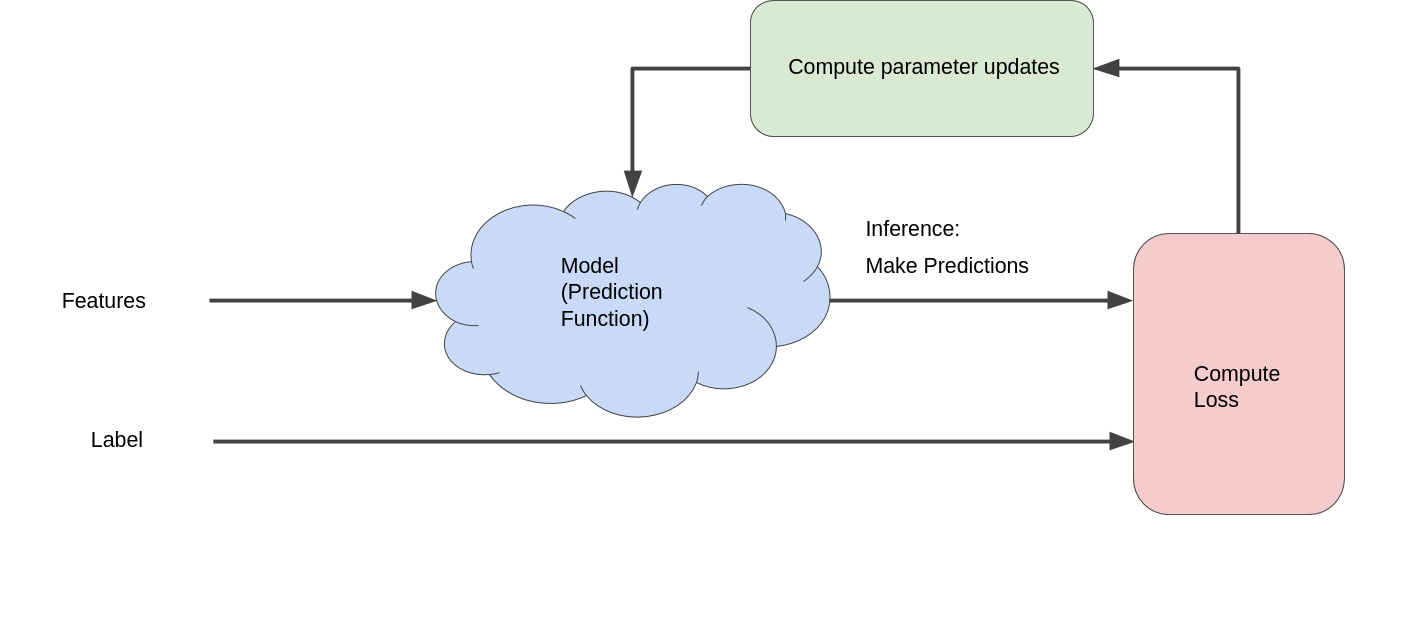
## Compiling a model

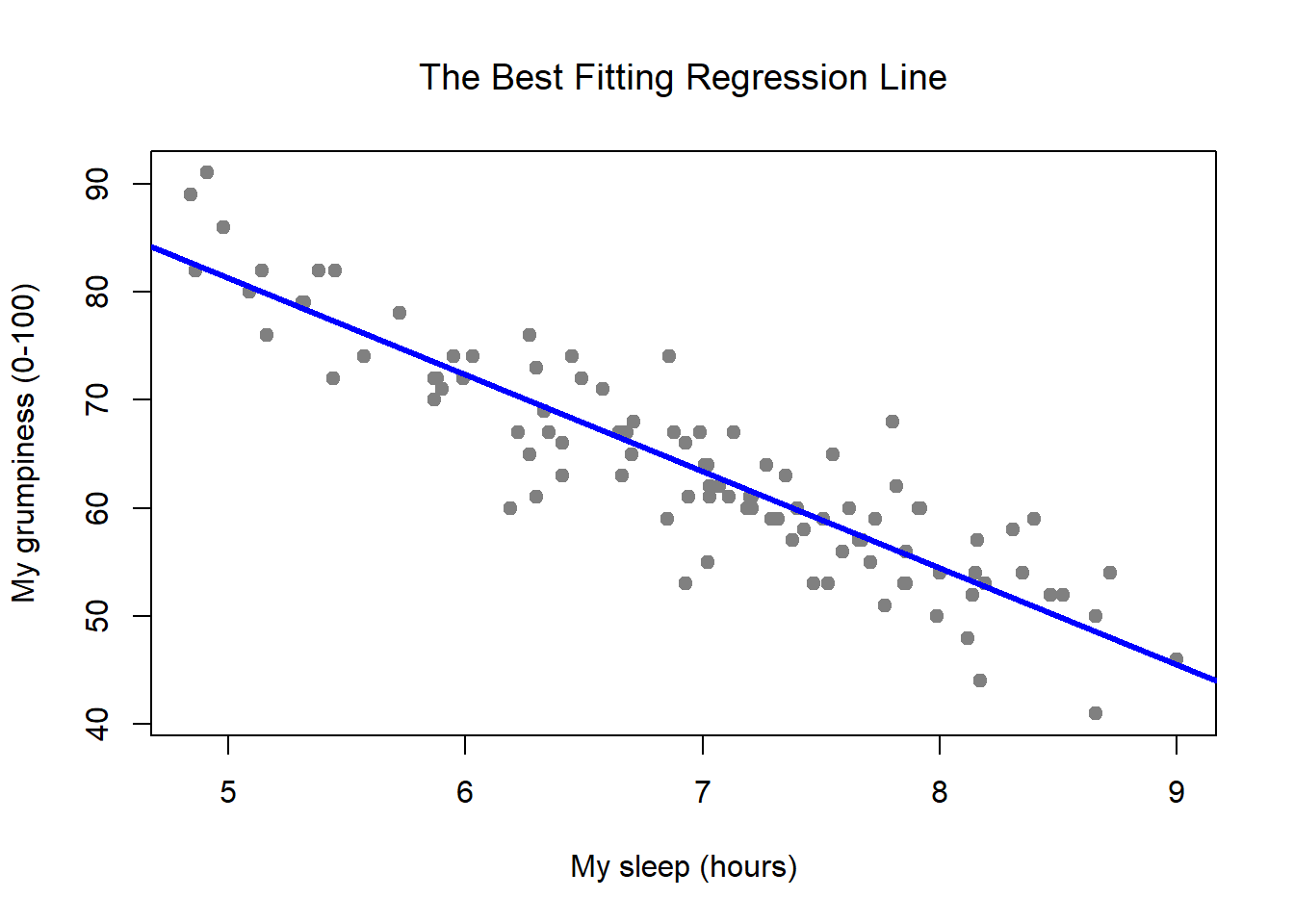
model.compile(optimizer='sgd', loss='mean\_squared\_error')

## Training the model

model.fit(features, labels, epochs=100)

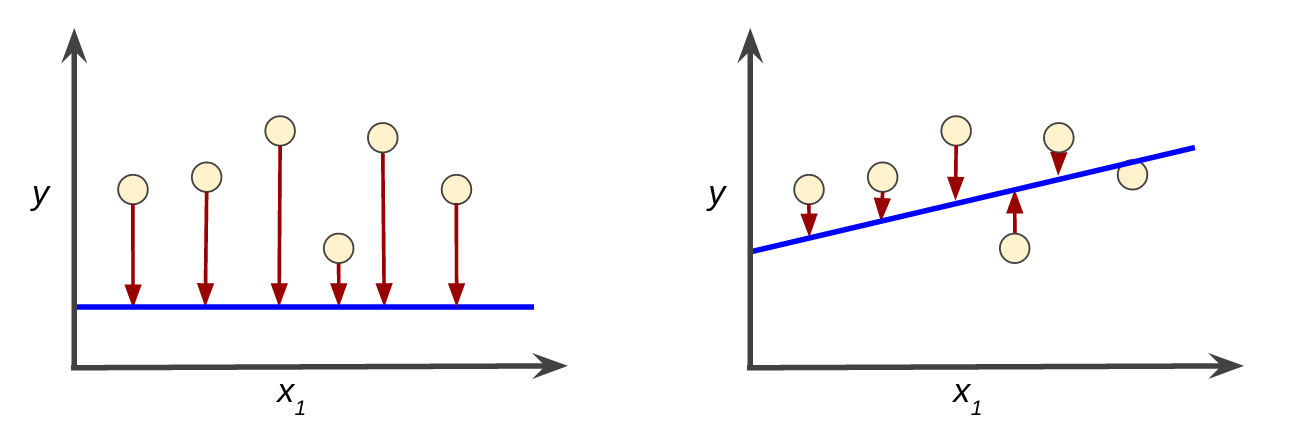
# How a model works



**Linear Regression**

* Finding the best possible curve to the given data.
* The data consists of features and the output (that we want the model to predict).
* For eg. in a grumpiness dataset, the features would be the total time I slept last night, the average sleep I had during the entire last week, etc.

**Loss**

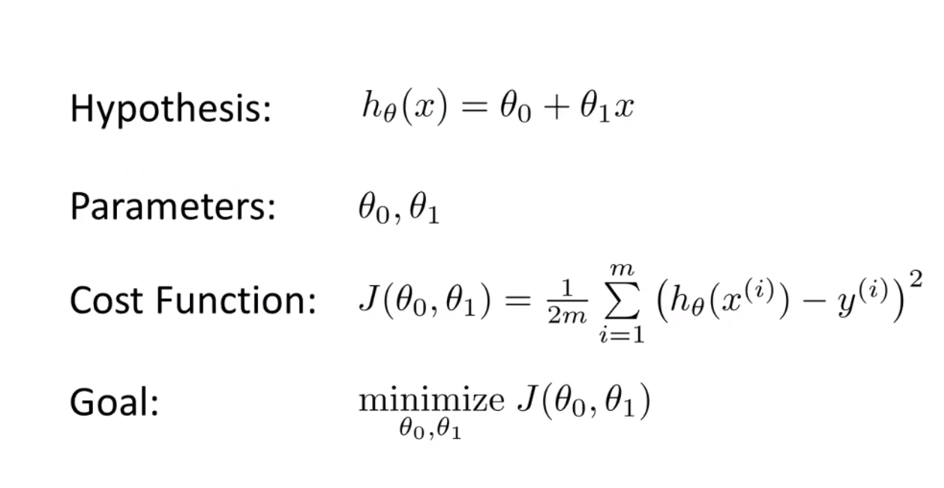
****

# 

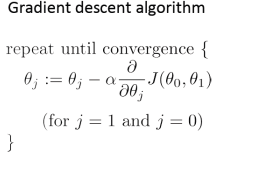
# Gradient Descent

An optimizer, something that tries to minimize a cost (or loss) function

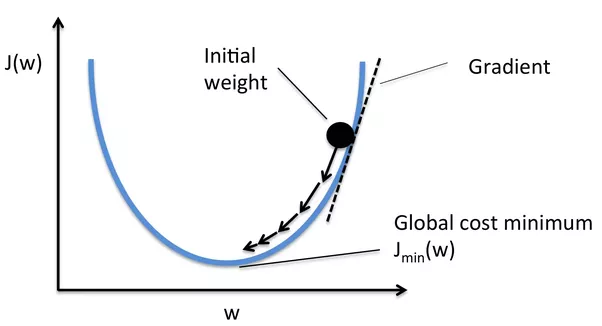
## The objective (Loss function considered in the example is mse)



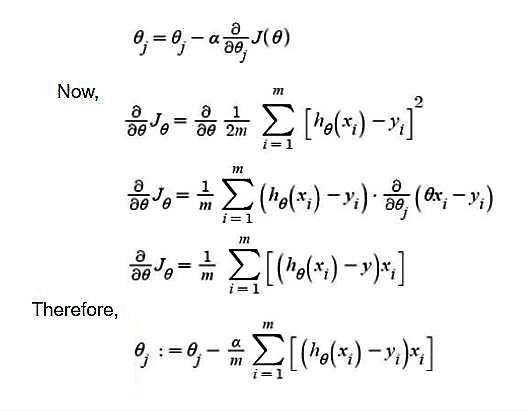
* **The Gradient Descent Algorithm**

****

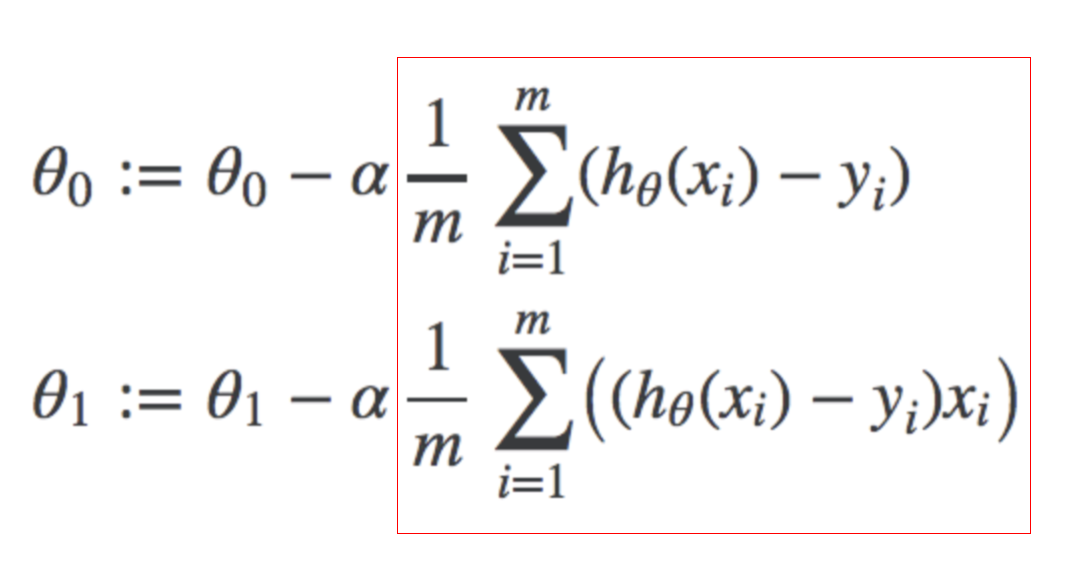
* **Gradient Descent gives us direction to move**



## Partial Derivative



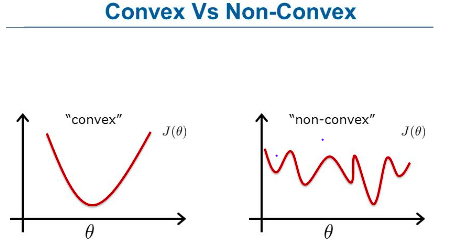
* **For our example**

****

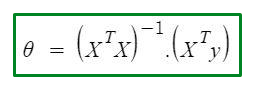
* **Gradient Descent Variations**

1. **Stochastic Gradient Descent**
2. **Mini-batch Gradient Descent**

## Does the starting point matter?

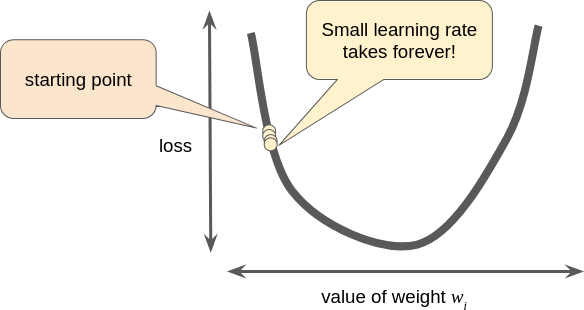
* ****

## Normal Equation

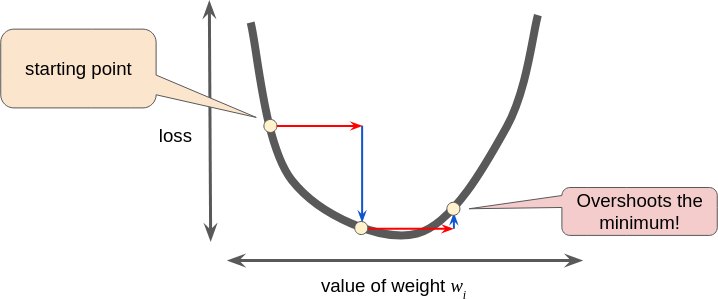
****

# Learning Rate (Hyperparameter)

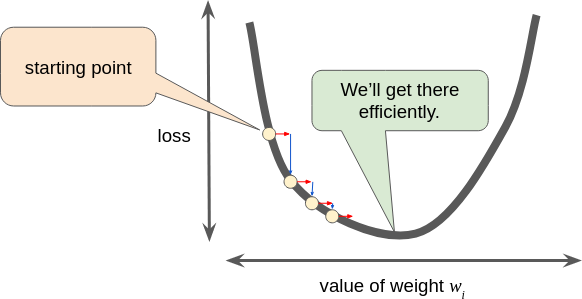
## Small Learning Rate

****

## Large Learning Rate

****

## Optimum Learning Rate

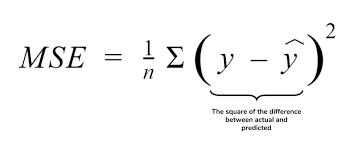
****

# WHAT IS A LOSS FUNCTION?

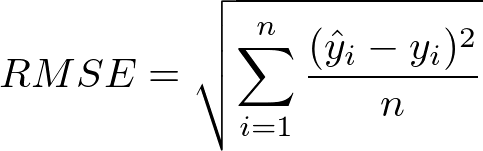
A way of evaluating the performance of your model. By comparing the output that my model generated to the correct (or true) output, I can have a better (quantifiable) estimate of how good or bad the model is performing.

# TYPES OF LOSS FUNCTION:

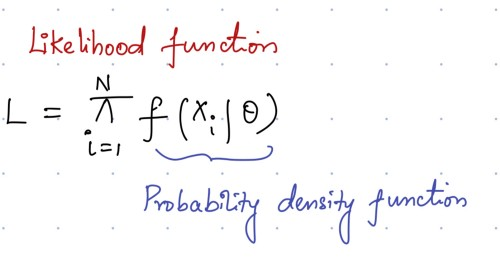
## 1] MEAN SQUARE ERROR (MSE)

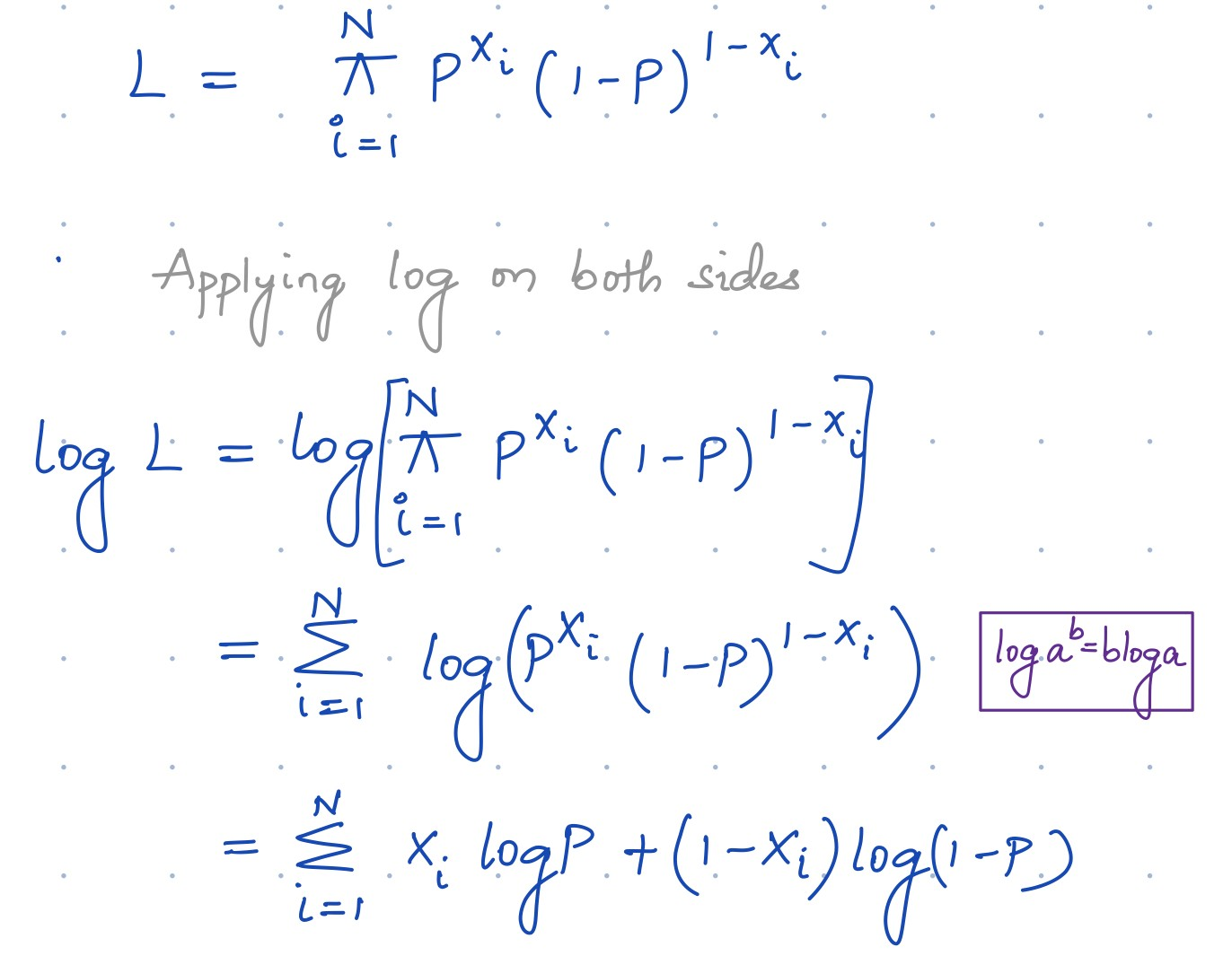


## 2] ROOT MEAN SQUARE ERROR (RMSE)



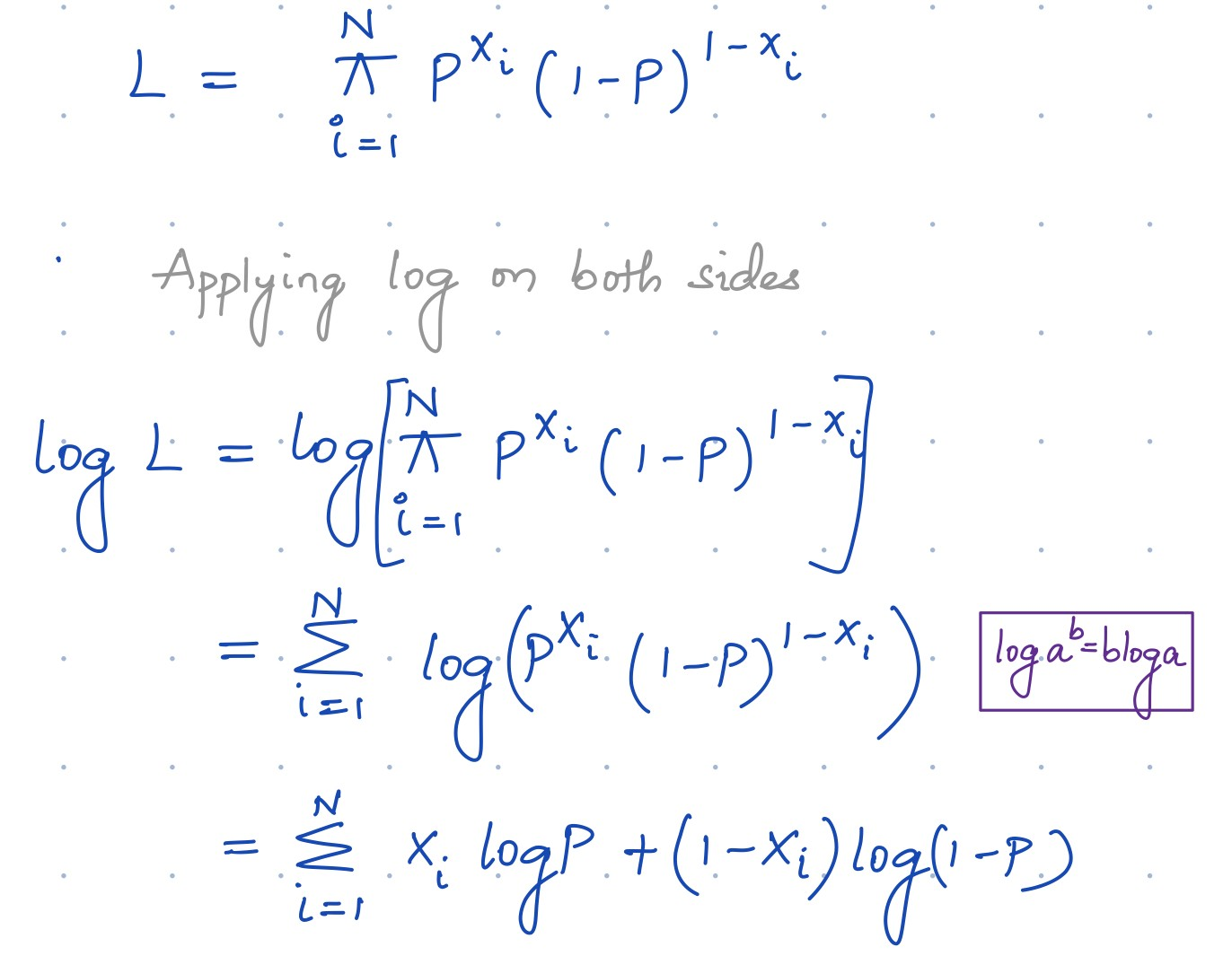
## 3] LIKELIHOOD FUNCTION

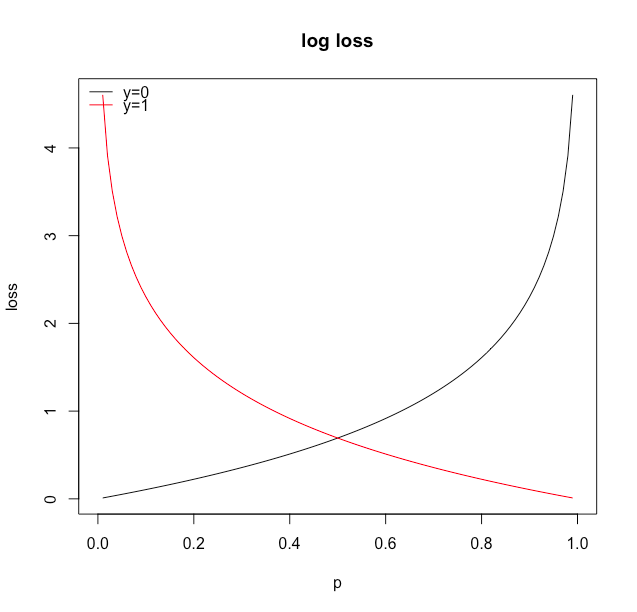


****

* **Xi - The truth (or correct) output.**
* **P - The probability that was generated by our model.**
* **If the correct output is 1, then we multiply with P**
* **If the correct output is 0, then we multiply with (1 - P)**

## 4] LOG LOSS (CROSS ENTROPY LOSS)

****



# 

# 

## Softmax Function

* Used to normalize the output of the model.
* So that the sum of the generated probabilities equals 1.

# 

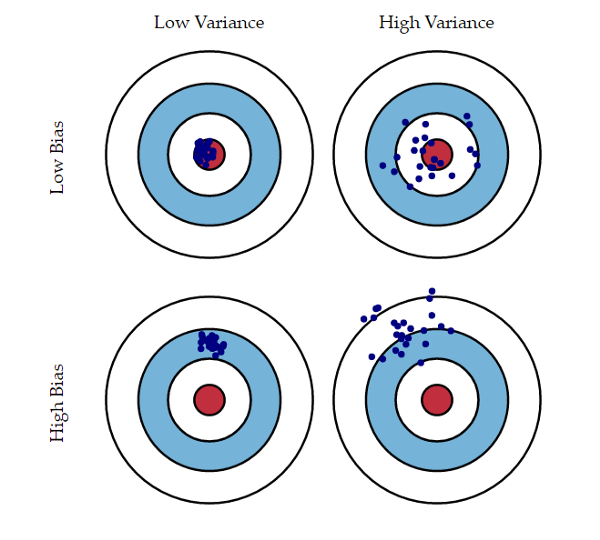
# How Do We Work With Datasets?

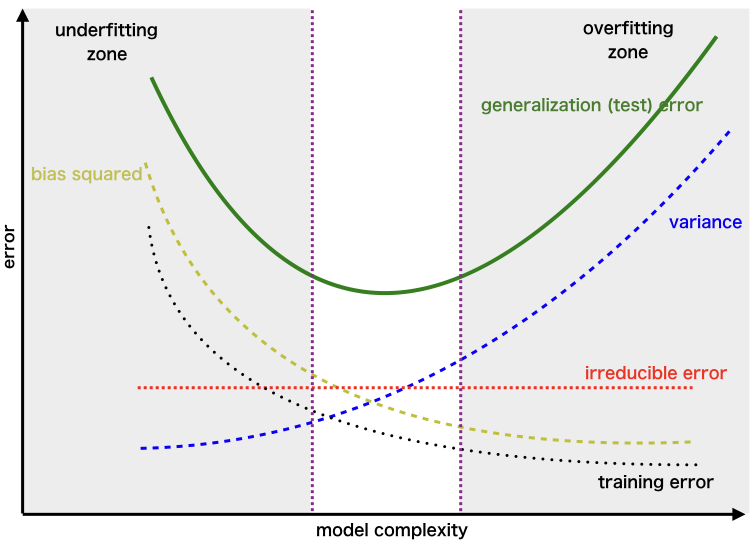
While working with a model, we divide the dataset into three parts:

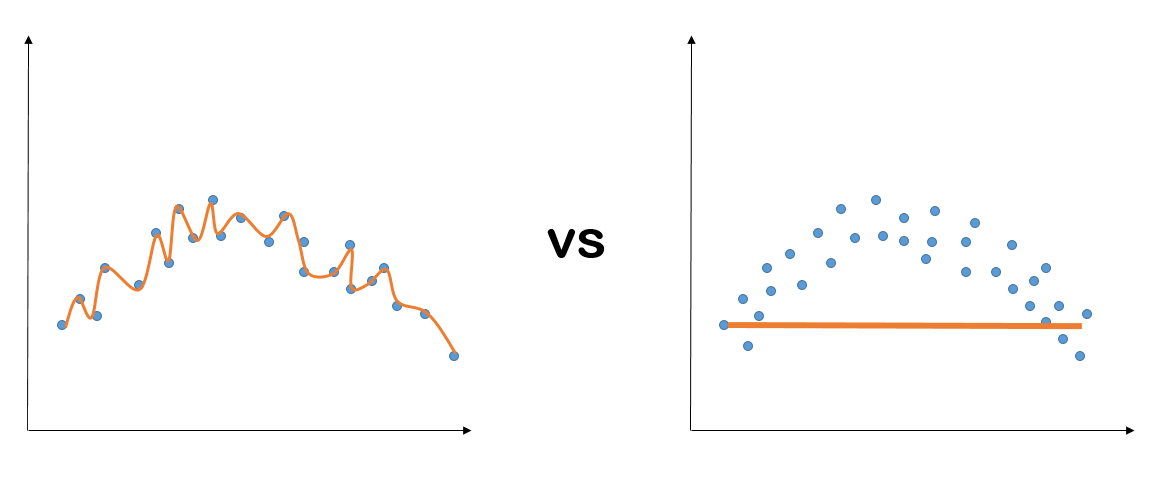
1. **Training data (70%)**
2. **Cross-validation data (15%)**
3. **Test data (15%)**

**Bias** - The preconception that the model already has wrt to the data.

**Variance** - The ability of the model to generate more variable kinds of data, if the variance is low this implies that the model can not fit well with varying data samples. But high variance leads to noise in the dataset.







A **High Variance model** vs a **High Bias model**

# TOPICS TO READ ABOUT

1. **What is supervised learning?**
   1. **Linear Regression**
   2. **Logistic Regression**
2. **Types of loss functions**
   1. **Mean Squared Error**
   2. **Root Mean Squared Error**
   3. **Likelihood Loss**
   4. **Log Loss (or Cross Entropy Loss)**
3. **Learning Rate**
   1. **Hessian Matrix**
4. **Bias vs Variance**
5. **Feature Scaling and Normalization**

# 

# WHAT IS A LOGISTIC CLASSIFIER?

# 