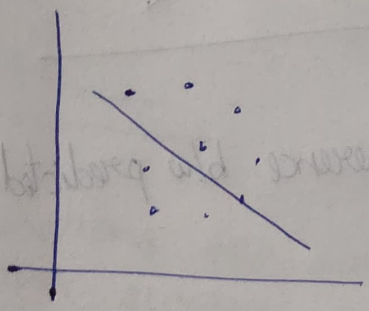
Machine Learning

Machine learning is a program that analyses data and learns to predict the outcome.

Dataset :- In the mind of a computer, a data set is any collection of data. It can be anything from an array to a complete database.

Regression :- It is used when you try to find relationship b/w variables, that relationship is used to predict the outcome of future events.

linear regression :- It is a statistical technique used to find the relationship between variables to draw a straight line through them.



In algebraic terms, the model would be defined as $y = mx + b$

In ML, we write equation for linear regression model as

$$y' = b + w_1 x_1$$

y' - the output

b - bias (y-intercept) is a parameter of model & is calculated during training.

w_1 - weight (m in algebra)

x_1 - input / feature.

- During training, the model calculates the weight and bias that produce the best model.

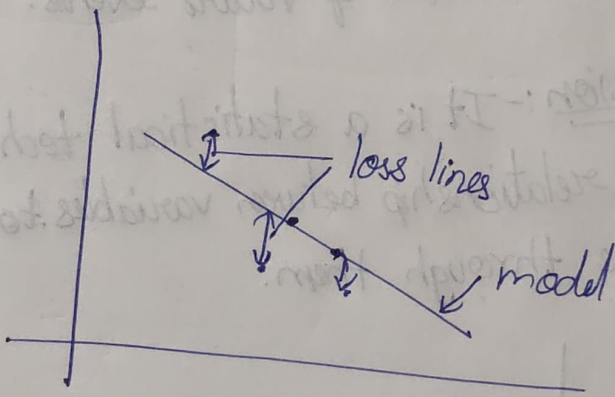
Models with multiple features :-

$$y' = b + w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots w_n x_n$$

Loss :- Loss is a numerical metric that describes how wrong a model's predictions are.

Loss measures the distance between the model's predictions & actual labels

The goal of training a model is to minimize the loss



Distance of loss is difference b/w predicted & actual value.

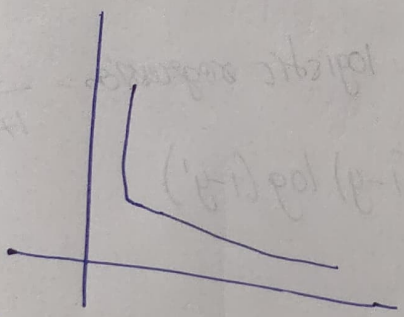
Types of loss :-

1) L_1 loss :- Σ |actual value - predicted value|
2) Mean absolute error :- $\frac{1}{N} \Sigma$ |actual value - predicted value|
3) L_2 loss :- Σ (actual value - predicted value)^2
4) Mean squared error :- $\frac{1}{N} \Sigma$ (actual value - predicted value)^2

Gradient descent is a mathematical technique that iteratively finds the weights and bias that produce the model with the lowest loss.

The model begins training with randomized weights and biases near zero and then repeats the following steps:

1) Calculate the loss with current weight & bias
2) Determine the direction to move the weights & bias that reduce loss.
3) Move the weight & bias values a small amount in the direction that reduces loss.
4) return to step 1 & repeat the process until the model can't reduce the loss any further.



— Typical loss curves

- The loss functions for linear models always produce convex surface

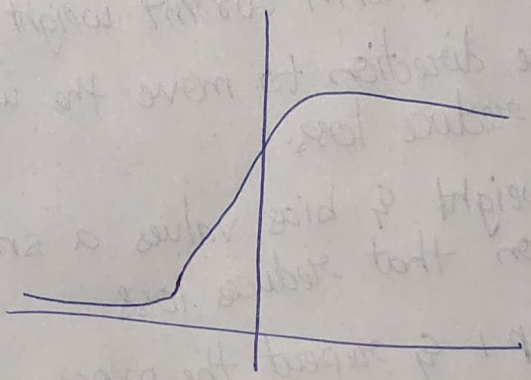
Hyper parameters :- These are variables that control different aspects of training. Three common hyperparameters are:

- learning rate
- batch size
- epochs.

Logistic regression:-

Logistic regression that is designed to predict the probability of a given outcome.

Sigmoid function $f(x) = \frac{1}{1+e^{-x}}$



If $z = b + w_1x_1 \cdots + w_nx_n$ logistic regression:- $\frac{1}{1+e^{-z}}$

log loss $= \sum_{(x,y)\in D} -y \log(y') - (1-y) \log(1-y')$