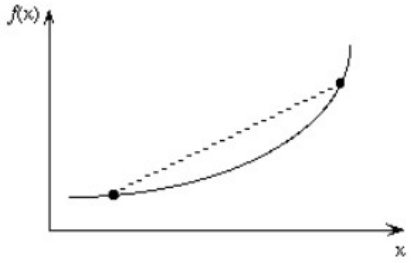
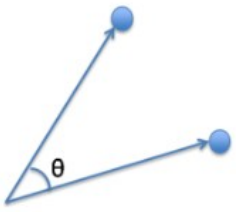
Convergence	Convergence refers to moving towards union or uniformity. An iterative algorithm is said to converge when as the iterations proceed the output gets closer and closer to a specific value.
Convex Function	A real value function is called convex if the line segment between any two points on the graph of the function lies above or on the graph.  Convex functions play an important role in many areas of mathematics. They are especially important in the study of optimization problems where they are distinguished by a number of convenient properties.
Correlation	Correlation is the ratio of covariance of two variables to a product of variance (of the variables). It takes a value between +1 and -1. An extreme value on both the side means they are strongly correlated with each other. A value of zero indicates a NIL correlation but not a non-dependence. You'll understand this clearly in one of the following answers. The most widely used correlation coefficient is Pearson Coefficient. Here is the mathematical formula to derive Pearson Coefficient. $$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2}\sqrt{\sum_i (y_i - \overline{y})^2}}$$
Cosine Similarity	Cosine Similarity is the cosine of the angle between 2 non-zero vectors. Two parallel vectors have a cosine similarity of 1 and two vectors at 90° have a cosine similarity of 0. Suppose we have two vectors A and B, cosine similarity of these vectors can be calculated by dividing the dot product of A and B with the product of the magnitude of the two vectors. $$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$ 

Cost Function	Cost function is used to define and measure the error of the model. The cost function is given by: $$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$ Here, • h(x) is the prediction • y is the actual value • m is the number of rows in the training set Let us understand it with an example: So let's say, you increase the size of a particular shop, where you predicted that the sales would be higher. But despite increasing the size, the sales in that shop did not increase that much. So the cost applied in increasing the size of the shop, gave you negative results. So, we need to minimize these costs. Therefore we make use of cost function to minimize the loss.
Covariance	Covariance is a measure of the joint variability of two random variables. It's similar to variance, but where variance tells you how a single variable varies, co variance tells you how two variables vary together. The formula for covariance is: $$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$ Where, *x* = the independent variable *y* = the dependent variable *n* = number of data points in the sample x bar = the mean of the independent variable *x* y bar = the mean of the dependent variable *y* A positive covariance means the variables are positively related, while a negative covariance means the variables are inversely related.
Cross Entropy	In information theory, the cross entropy between two probability distributions and over the same underlying set of events measures the average number of bits needed to identify an event drawn from the set, if a coding scheme is used that is optimized for an "unnatural" probability distribution , rather than the "true". Cross entropy can be used to define the loss function in machine learning and optimization.
Cross Validation	Cross Validation is a technique which involves reserving a particular sample of a dataset which is not used to train the model. Later, the model is tested on this sample to evaluate the performance. There are various methods of performing cross validation such as: • Leave one out cross validation (LOOCV) • k-fold cross validation • Stratified k-fold cross validation • Adversarial validation