Optimisation Algorithm Techniques :

1. Gradient Descent (We used batch gradient descent, batch Stochastic gradient descent and simple gradient descent algorithms to solve optimization problems .)
2. Back propagation
3. Convergence theorem
4. Learning rate
5. Initialization
6. Normalization
7. Batch SGD with Momentum
8. Nesterov Accelerated Gradient (NAG)
9. AdaGrad
10. Adadelta and RMSProp
11. Adam (Adaptive Moment Estimation)

Loss Functions :

1. Cross – Entropy
2. Hinge
3. Huber
4. Kullback-Leibler
5. MAE(L1)
6. MSE(L2)

Cost Functions :

A cost function is a measure of error between what value your model predicts and what the value actually is. The cost function has its own curve and its own gradients. The slope of this curve tells us how to update our parameters to make the model more accurate.

An activation function transforms the shape/representation of the data going into it.

**Loss function** is a method of evaluating “how well your algorithm models your dataset”. If your predictions are totally off, your loss function will output a higher number. If they’re pretty good, it’ll output a lower number. As you tune your algorithm to try and improve your model, your loss function will tell you if you’re improving or not. ‘Loss’ helps us to understand how much the predicted value differ from actual value

**Gradient Descent**

* Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient.
* In machine learning, we use gradient descent to update the [parameters](https://ml-cheatsheet.readthedocs.io/en/latest/glossary.html#glossary-parameters) of our model.
* Parameters refer to coefficients in [Linear Regression](https://ml-cheatsheet.readthedocs.io/en/latest/linear_regression.html) and [weights](https://ml-cheatsheet.readthedocs.io/en/latest/nn_concepts.html#nn-weights) in neural networks.

Gradient descent is an optimization algorithm used to find the values of parameters (coefficients) of a function (f) that minimizes a cost function (cost).

**Adam :**

* **Adam** is a replacement optimization algorithm for stochastic gradient descent for training **deep learning** models.
* **Adam** combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems.

