Write a paragraph on loss functions

- The loss function is the function that computes the distance between the current output of the algorithm and the expected output. It's a method to evaluate how your algorithm models the data.
- Loss functions measure how far an estimated value is from its true value.Loss functions are not fixed, they change depending on the task in hand and the goal to be met.
- A loss function takes a theoretical proposition to a practical one. Building a highly accurate predictor requires constant iteration of the problem through questioning, modeling the problem with the chosen approach and testing.
- The only criteria by which a statistical model is scrutinized is its performance - how accurate the model's decisions are. This calls for a way to measure how far a particular iteration of the model is from the actual values. This is where loss functions come into play.

I have written more about it here - https://docs.google.com/document/d/12pM8Gea6Ke4DIEnFtPa0PXNaIJD1GNjgoh4Z9GEWFYU/edit?usp=sharing

Write a paragraph on different optimizers

Adam Optimizer

Adam stands for **adaptive moment estimation**, which is another way of using past gradients to calculate current gradients, for the deep mathematical explanation you can read its official paper(Kingma & Ba, 2014) here, Adam utilizes the concept of momentum by adding fractions of previous gradients to the current one, it is practically accepted in many projects during training neural nets.

SGD Optimizer

Stochastic gradient descent(SGD) optimization algorithm in contrast performs a parameter update for each training example. SGD performs redundant computations for bigger datasets, as it recomputes gradients for the same example before each parameter update.

NAdam Optimizer

NAdam optimizer is an acronym for **Nesterov** and **Adam optimizer**. Its official research paper was published in 2015 here, now this Nesterov component is way more efficient than its previous implementations. Nadam used Nesterov to update the gradient.

AdaMax Optimizer Class

As the name suggests AdaMax is an adaption of Adam optimizer, by the same researchers who wrote the Adam algorithm

Adadelta(adaptive delta) Optimizer

Now like the **RMSprop** optimizer, **Adadelta**(Read paper: Zeiler, 2012) is another more improved optimization algorithm, here delta refers to the difference between the current weight and the newly updated weight. Adadelta removed the use of the learning rate parameter completely and replaced it with an exponential moving average of squared deltas.