**Types of Gradients in Keras:**

Great videos of Andrew Ng on optimizer is given below

[**https://www.youtube.com/watch?v=4qJaSmvhxi8**](https://www.youtube.com/watch?v=4qJaSmvhxi8)

* Mini batch Gradient descent
* SGD
* RMSprop
* Adam
* Adadelta
* Adagrad
* Adamax
* Nadam
* Ftrl

1. **SGD:**

Keras provides the SGD class that implements the stochastic gradient descent optimizer with a learning rate and momentum. The default learning rate is 0.01 and no momentum is used by default.

1. **RMSprop:**

RMSprop is another gradient based optimization technique. RMSprop uses an adaptive learning rate instead of treating the learning rate as a [hyperparameter](https://deepai.org/machine-learning-glossary-and-terms/hyperparameter). This means that the learning rate changes over time.

1. **Adam:**

Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments. Adam can be looked at as a combination of RMSprop and Stochastic Gradient Descent with momentum. It uses the squared gradients to scale the learning rate like RMSprop and it takes advantage of momentum by using moving average of the gradient instead of gradient itself like SGD with momentum.

Multihead Network

<https://debuggercafe.com/multi-head-deep-learning-models-for-multi-label-classification/>