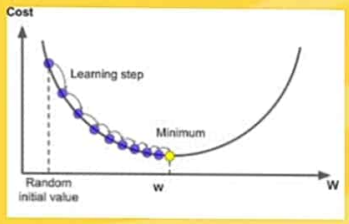
GRADIENT DESCENT

- Consider Equation 2x^2
- Random initialization of x
- Learning_rate To determine the step size
- Gradient = dy/dx = d/dx(2x^2)
  = 4x

- repeat until convergence

{
w = w - (learning_rate * (dJ/dw))
b = b - (learning_rate * (dJ/db))
}



1. If the learning rate is too large, gradient descent will miss the local minimum and fluctuate at different points uncontrollably
2. If the learning rate is too small, the algorithm will take too long to reach the minimum and eat up your computer's power.

BATCH GRADIENT DESCENT

- Processes all the training examples for each iteration of gradient descent
- But if the number of training examples is large, then batch gradient descent is computationally very expensive

STOCHASTIC GRADIENT DESCENT

- **Processes 1 training example per iteration**
- **Hence, the parameters are being updated even after one iteration in which only a single example has been processed**
- **Number of iterations will be quite large**

- One approach to the problem of stochastic gradient descent not being able to settle at a minimum is to use something known as a **learning schedule, seek to adjust the learning rate during training by reducing the learning rate according to a pre-defined schedule**

- this gradually reduces the learning rate

- If the learning rate is reduced too quickly, then the algorithm may get stuck at local minima

- If the learning rate is reduced too slowly, you may jump around the minimum for a long time and still not get optimal parameter

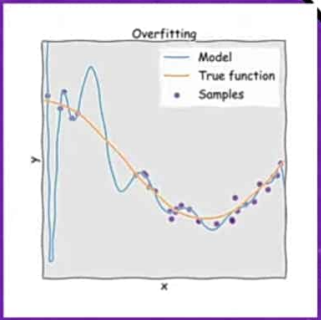
Mini Batch Gradient Descent

- Instead of calculating the gradient on the whole dataset, or random examples of the dataset, it calculates them on small subsets of the dataset, commonly referred to as mini-batches
- Works faster than both batch gradient descent and stochastic gradient descent
- it works for larger training examples and that too with lesser number of iterations.

Deep Learning Terminology - 1

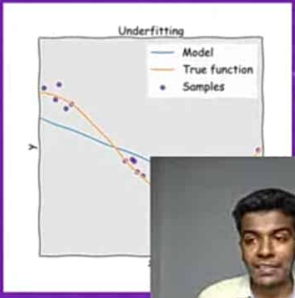
Overfitting

- Good performance on the training data, poor generalization to other data.



Underfitting

- Poor performance on the training data and poor generalization to other data



Deep Learning Terminology - 2

Epochs

- Single training iteration of all batches in both forward and back propagation
- 1 epoch is a single forward and backward pass of the entire input data

Deep Learning Terminology - 3

Dropout

- Dropout is a regularization technique which prevents over-fitting of the network
- During training a certain number of neurons in the hidden layer is randomly dropped