**Neural Networks**

**Course 2:**

**Hyper Parameters Tuning**

**Week 2-3**

**Submitted to Dr. Omar Nasr**

**Prepared by: Mona Mamdouh**

# **Mini-Batch gradient decent**

The data is divided into mini-batches as follows

Batch#3

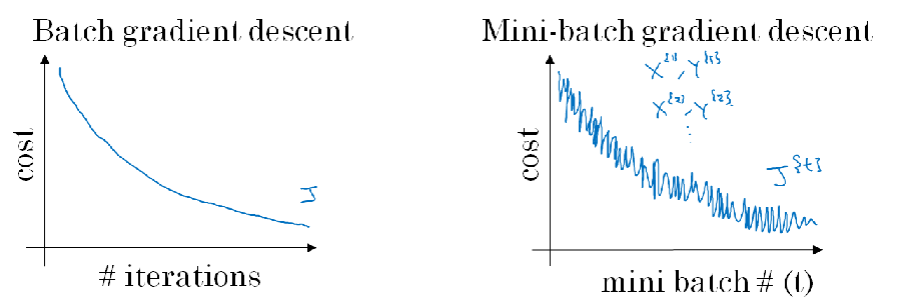
Batch#m

Batch#2

Batch#1

…………………..

The network is trained by each batch and take one gradient step, the cost function in mini batch is noisy as the cost is measured on a mini batch.



* Choose the mini batch size is according to :
  + Mini-batch size = m (returns to batch gradient decent)
  + Mini-batch size = 1 (stochastic gradient decent 'very noisy')
  + Practically the used number is in between to make progress without waiting all the data to be processed.
* To remove the noise in the cost function in case of mini batch we use the momentum average technique.

# **The momentum gradient decent**

* In iteration t we compute dw, wb, on the current mini-batch
* Then we compute

Where a new hyper parameter is usually set to 0.9,

And is initialized to zero

# **RMS prop**

* In iteration t we compute dw, wb, on the current mini-batch
* Then we compute

Where a new hyper parameter,

# **Adam optimization algorithm**

* In iteration t we compute dw, wb, on the current mini-batch
* Then we initialize

Where a new hyper parameter,