Different types of Gradient Descent

① Batch gradient Descent

- input → entire traiging dataset
- we will use all dataset to Compute the gradient at each itteration and then Back to adjust weights
- it is good for small training set
- But too expensive for large datasets
- B. it require to load dataset into memory

② Stochastic gradient descent

- input → one training example at a time
- it will Randomly pick one sample for each step or itteration and use this sample to Calculate the derivatives
- it used For big data it is useful if we have redundancies in data we don't want too much Cmputation

③ MiniBatch gradient Desce

- input → Multiple traini example at a time but less than entire dataset
- ~~Mini~~ we will use a batch of randomly picked samples
- Mini Batches will used to Cmpute gradient
- it Cmbine benefit of Batch and stochastic

④ Momentum Gradient Descent

- introduce a velocity term that accumulates the past gradient
- Faster Cnvergence
- Better For handeling noise
- Reduce oscillation
- Momentum ↑↑ overshoot

⑤ Nesterov Accelerated gradient

- Enhancement over Momentum based Method Calculate gradient at future position
- ~~prov~~ Looking ahead → more accurate
- Faster Cnvergance than standard Momentum
- More Cmplex