

---

# How does Batch Normalization Help Optimization?

---

Abraham Jose  
ID :5068109, CAP6614  
abraham@knights.ucf.edu

## 1 Summary

It is systematic empirical study on Batch Normalization and how Batch Normalization helps us to achieve a smooth landscape for optimization of the loss. The study demystifies the fact that the Batch Normalization doesn't correlate with Internal Co variance Shift, which is a popular belief and understanding of how Batch Normalization works.

### 1.1 Strengths of the proposal

1. Proves Explicitly that Internal Co-variance Shift has no effect on the optimization or the training accuracy performance.
2. Theoretically proves how Batch Normalization stabilizes and smooths the optimization landscape and studies the effect of batch normalization on the optimization landscape including how it fixes first order moments of activation.
3. Compares Batch Normalization with the other optimization landscape smoothing techniques.

### 1.2 Weaknesses of the proposal

1. The possible ways to increase the efficiency of Batch Normalization based on the understanding of how Batch Normalization works.
2. Effects of Batch Normalization on the initialization of the network and the type of optimization is not discussed.
3. In proving how Batch Normalization usually helps in better generalization.

### 1.3 Results

A systematic investigation of Batch Normalization in a neural network and how Batch Normalization helps us to achieve a smooth optimization landscape.

## 2 Discussion

### 2.1 Future Work

A study on how we can further smooth the optimization landscape with the knowledge we acquired on how the Batch Normalization works for faster convergence to the optimal solution.

Effect of Internal Co variant shift in a deep neural network and how it affects the training of the deep neural network.

### 2.2 Q & A

- If the Internal Co variant Shift has no impact in the optimization landscape, why is that initialization important, given it helps to achieve better training accuracy in less iterations.
- Since Batch Normalization affects the optimization landscape, does Batch Normalization is dependent on the type of optimization that we choose?