## Abstract

Deep neural nets with a large number of parameters are very powerful machine learning systems. However, overfitting is a serious problem in such networks. Large networks are also slow to use, making it difficult to deal with overfitting by combining the predictions of many different large neural nets at test time. Dropout is a technique for addressing this problem. The key idea is to randomly drop units (along with their connections) from the neural network during training. This prevents units from co-adapting too much. During training, dropout samples from an exponential number of different "thinned" networks. At test time, it is easy to approximate the effect of averaging the predictions of all these thinned networks by simply using a single unthinned network that has smaller weights. This significantly reduces overfitting and gives major improvements over other regularization methods. We show that dropout improves the performance of neural networks on supervised learning tasks in vision, speech recognition, document classification and computational biology, obtaining state-of-the-art results on many benchmark data sets.

Keywords: neural networks, regularization, model combination, deep learning

- 가능한 Parameter 조합에 따른 Prediction 값의 가중평균 (Model Combination)

- Model Combination을 사용하면 성능은 향상되지만, DNN의 규모와 연산량 증가

- Dropout의 효과는 여러 개의 "Thinned Network"를 학습시키는 것과 같음

- Thinned Network == Dropout 과정에서 생략되지 않은 뉴런들로만 구성

(a) Standard Neural Net   (b) After applying dropout.

(a) At training time   (b) At test time
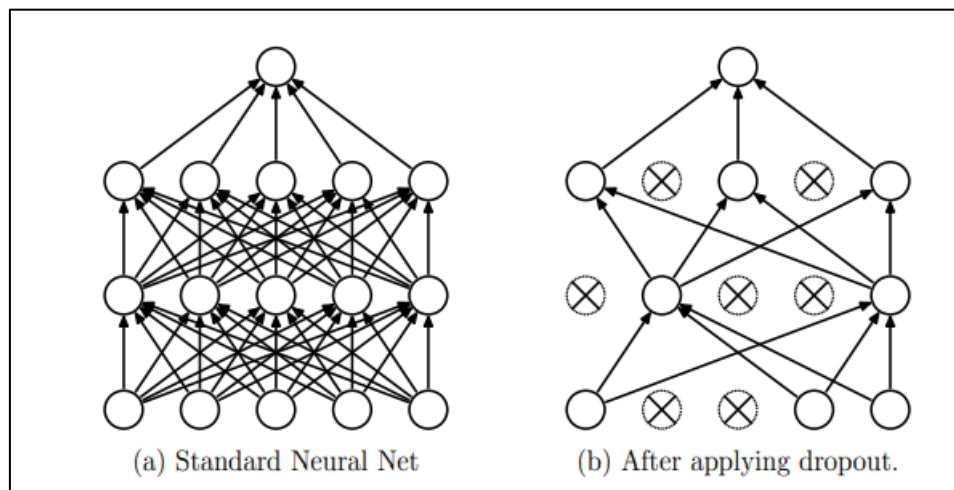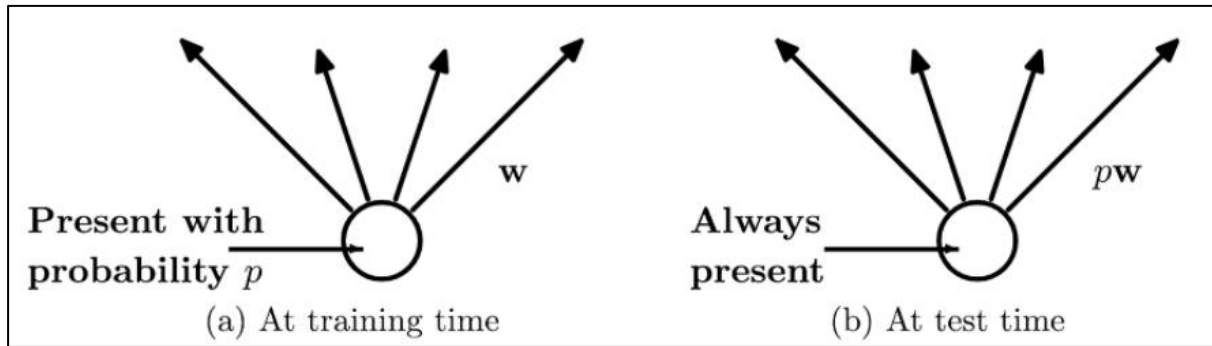
- 조합 가능한 Parameter로 이루어진 Thinned Network들 각각의 Prediction의 가중평균보다는

- p * w를 통해 approximate averaging method를 제안

making the presence of any particular hidden unit unreliable. This technique was found to improve the performance of neural nets in a wide variety of application domains including object classification, digit recognition, speech recognition, document classification and analysis of computational biology data. This suggests that dropout is a general technique and is not specific to any domain. Methods that use dropout achieve state-of-the-art results on SVHN, ImageNet, CIFAR-100 and MNIST. Dropout considerably improved the performance of standard neural nets on other data sets as well.

- General Technique, not specific to any domain

- 실제로 모든 DNN에 사용됨