The Korean Journal of Applied Statistics (2019) DOI: http://dx.doi.org/10.5351/KJAS.2019.32.6.899
32(6), 899–908

Hybrid dropout

Chongsun Park[a,1] · MyeongGyu Lee[a]

[a]Department of Statistics, Sungkyunkwan University

(Received October 2, 2019; Revised November 1, 2019; Accepted November 5, 2019)

Abstract

Massive in-depth neural networks with numerous parameters are powerful machine learning methods, but they have overfitting problems due to the excessive flexibility of the models. Dropout is one methods to overcome the problem of oversized neural networks. It is also an effective method that randomly drops input and hidden nodes from the neural network during training. Every sample is fed to a thinned network from an exponential number of different networks. In this study, instead of feeding one sample for each thinned network, two or more samples are used in fitting for one thinned network known as a Hybrid Dropout. Simulation results using real data show that the new method improves the stability of estimates and reduces the minimum error for the verification data.

Keywords: deep neural network, dropout, co-adaptation, overfitting, hybrid method

1. 서론

수많은 모수들을 포함하는 심층신경망모형(deep neural network model)은 강력한 기계학습 시스템을 구현할 수 있다. 그러나 모형의 과도한 융통성으로 인하여 훈련자료(training dataset)에 대한 적합도가 높음에도 훈련자료 크기의 한계로 인한 공적응(co-adaptation)이 발생하게 되며 결과적으로 훈련자료에 대한 오차에 비하여 검증자료(testing dataset)에 대한 예측력이 상대적으로 떨어지는 과적합(overfitting)문제를 내포하고 있다.

이러한 과적합 문제를 해결하기 위하여 사용할 수 있는 기법들은 다양하게 존재한다. 가장 간단한 방법으로는 과적합이 일어나지 않을 정도의 단순한 모형을 사용하는 방법을 생각할 수 있다. 또한 신경망모형의 경우 조기중단(early stopping)을 이용하여 과적합이 발생하기 전에 학습을 중단하는 방법을 적용할 수 있다. 다른 시도로 L_1과 L_2 벌점함수를 이용한 일반화(regularization)와 정규혼합분포를 이용하여 벌점함수를 모형화하는 방법 (Nowlan과 Hinton, 1992)을 들 수 있다. 미니배치(mini-batch) 단위로 표준화를 하는 배치표준화방법(batch normalization) (Ioffe와 Szegedy, 2015) 또한 일반화의 효과가 있어 아래에 소개할 드롭아웃(dropout)방법을 대체할 수 있는 것으로 알려져 있다. 일반화의 다른 방법으로 다양한 모형들의 평균모형을 이용하여 예측하는 방법 (Xiong 등, 2011; Salakhutdinov와 Mnih, 2008)이 효과적이나 신경망의 크기가 상대적으로 작은 경우에만 적용할 수 있어 적용범위가 한정되어 있는 단점이 있다.

[1]Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2 sungkyunkwan-Ro, Jongno-Gu, Seoul 03063, Korea. E-mail: cspark@skku.edu

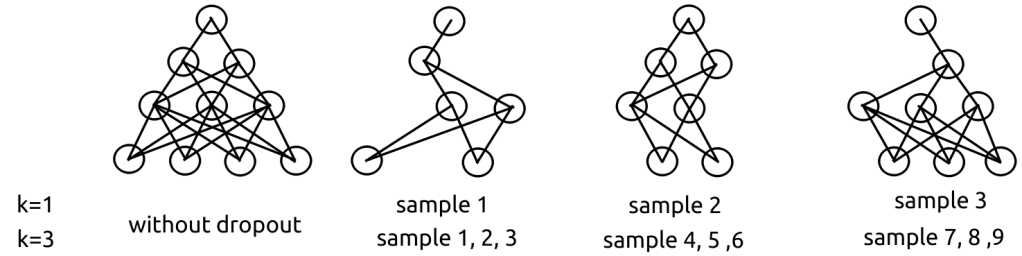


Figure 2.1. Small example of neural network with and without dropout when $k = 1$ and $k = 3$.

Hinton 등 (2012)에 의하여 소개된 드롭아웃방법 (Srivastava, 2013; Srivastava 등, 2014)은 각각의 표본마다 다른 모형을 학습시켜 융통성이 큰 모형이 갖는 단점 중 하나인 공적응을 방지하기 위하여 고안되었다. 과도한 융통성을 갖는 모형은 적절하게 행동하지 못하는 노드가 있는 경우에도 다른 노드(node)들의 도움으로 참의 모형과 비슷한 수준의 예측이 가능하게 되는데 이를 공적응이라고 한다. 이러한 현상의 발생을 방지하는 방법으로 모든 표본의 학습시에 동일한 모형을 사용하는 대신 표본이 바뀔 때마다 다른 모형을 적합하도록 하는 것이 드롭아웃이다. 드롭아웃 방법은 모든 표본이 동일한 모형을 사용하지 못하도록 각각의 표본이 사용할 모형을 선택할 때 모형에 포함된 노드 또는 활성함수(activation function)들을 무작위로 선택하게 된다. 드롭아웃 방법의 일반화 방법으로 무작위로 선택하는 대상을 노드가 아닌 가중치(weight)로 두고 선택된 가중치를 0으로 두는 DropConnect 방법 (Wan 등, 2013)도 있다.

완전연결신경망(fully-connected neural network)에 드롭아웃을 적용하게되면 훈련 초기에는 에포크(epoch)가 진행됨에 따라 오차의 변동이 드롭아웃을 적용하지 않은 경우에 비하여 작으나 이 현상이 오차가 특정 수준에 수렴하여도 계속되는 현상이 나타난다(아래의 MNIST 자료에 대한 Figure 3.1 참고). 본 연구에서는 모형이 수렴한 상태에서도 계속되는 오차의 변동성을 완화시키기 위하여 하나의 표본에 하나의 모형을 적합시키는 대신 하나의 모형에 둘 이상의 표본을 적합하여 모수 추정치들의 안정성을 높이는 방안을 고려하였다. 잘 알려진 MNIST와 CIFAR-10자료에 대하여 적용한 결과 새로운 방법을 적용하는 경우 오차의 변동성을 완화시킬 뿐만 아니라 모형에 대한 검증오차도 더 작게 나타나는 결과를 가져왔다.

2. 하이브리드(hibrid) 드롭아웃

Figure 2.1의 왼쪽 첫 번째 모형을 예로 완전연결신경망 모형의 구조를 살펴보자. 입력변수는 4개이고 입력층(input layer)에 연결된 2개의 은닉층(hidden layer)에는 각각 3개와 2개의 노드가 포함되어 있으며 마지막으로 최종 결과층(output layer)은 하나의 노드를 가지고 있다. 크기가 4인 입력벡터를 $\mathbf{x}$, 은닉층의 노드들을 $\mathbf{h}_{ij}$라 두면 첫 번째 은닉층의 가장 오른쪽 노드는 가중치 벡터를 $\mathbf{w_{11}}$, 활성함수가 정류선형유닛(rectified linear unit; ReLU)인 경우 $\mathbf{h}_{11} = \text{ReLU}(\mathbf{w_{11}}^T\mathbf{x} + b_{11})$로 나타낼 수 있다. 여기서, $\mathbf{w}_{11}$은 크기가 4인 모수벡터로 가중치가 되며 b_{11}은 스칼라로 바이어스가 된다.

2.1. 드롭아웃

앞에서 설명된 바와 같이 드롭아웃은 각각의 노드들이 다른 노드들의 도움을 받지 않고 스스로 정확한 판단 또는 역할을 할 수 있도록 고안되었다. 공적응 현상은 표본의 수는 한정되어 있고 이에 비하여 모형의 크기가 상대적으로 큰 경우 특정 노드가 올바르게 작동하지 못하여도 다른 노드들의 도움을

받아 훈련자료에 대한 오차는 작으나 검증자료에 대하여 적절한 예측을 하지 못할 때 발생하게 된다. Strivastava 등 (2014)은 이러한 공적응 현상을 방지하기 위하여 각각의 표본이 훈련에 사용하는 모형들이 서로 상이하도록 모형의 노드들을 무작위로 제거한 후 훈련하는 방법을 사용하였다. 드롭아웃 방법은 초기에 적당한 크기의 신경망을 선택하고 각각의 표본에 대한 훈련모형에서 임의로 노드들을 선택할 확률 p가 주어지면 입력층과 은닉층에 포함된 각각의 노드에 대하여 성공확률이 p인 베르누이 분포에서 난수를 추출하고 이 값이 1이면 해당 노드를 모형에 포함한다. Figure 2.1에는 입력 변수가 4개이고 두 개의 은닉층을 가지고 있으며 첫 번째 은닉층은 3개의 노드를 그리고 두 번째 은닉층은 2개의 노드를 가지고 있고 하나의 출력노드를 갖는 완전연결신경망 모형과 입력층과 은닉층의 노드들을 확률적으로 선택하여 처음 3개의 표본들의 훈련에 사용될 모형들이 포함되어 있다. 예제로 사용한 모형의 경우 크기가 작아 무작위로 노드들을 선택한 모형 중 입력에서 출력으로의 연결이 존재하지 않아 학습이 불가능한 경우가 어렵지 않게 발생할 수 있으나 실제 분석에 사용되는 모형들은 크기가 매우 커 이러한 현상이 발생할 확률은 매우 적다.

드롭아웃의 원리는 전체 훈련자료에 대하여 다수의 모형을 적합하고 그 결과를 통합하여 예측하는 평균모형 방법 또는 앙상블 방법과 유사하지만 표본마다 다른 모형을 사용하는 점이 다르다고 할 수 있다. 일반적인 앙상블 알고리즘들은 앞에서도 언급한 바와 같이 모형이 큰 경우 예측을 위하여 많은 계산이 필요하다는 단점이 있으나 드롭아웃에서는 검증시 모든 노드들이 포함된 완전모형을 사용하여 크기에 따른 연산요구량의 증가를 막을 수 있다.

검증시 사용되는 완전모형은 각각의 노드에 연결된 가중치들의 값을 그대로 사용하는 대신 훈련과정에서 모형에서 선택된 비율 p를 곱한 값을 사용하게 된다. 예를 들어 훈련에서 사용된 모형이 완전모형의 50%에 해당하는 노드만을 포함하였다면 동일한 표본에 대한 예측값을 적절하게 추정하려면 완전모형의 경우보다 모형의 가중치가 평균적으로 2배 큰 값을 가져야 하므로 각 노드가 훈련에 사용된 비율 p를 곱한 가중치를 가지는 완전모형을 검증에 사용한다. 따라서 일반적인 앙상블 알고리즘들의 경우 고려하는 모형의 수가 많은 경우 예측에 많은 연산이 필요하나 드롭아웃 방법의 경우는 모형의 크기가 큰 경우에도 예측에 필요한 연산에 대한 부담이 매우 적다고 할 수 있다.

Figure 3.1에 모의실험에서 사용한 MNIST자료에 드롭아웃을 사용하지 않은 모형과 사용한 모형을 적용한 결과 훈련자료와 검증자료의 오차그림이 포함되어 있다. 사용된 모형은 완전연결신경망 모형이며 2개의 은닉층과 각각의 은닉층에 2,048개의 노드를 포함하고 있다. 드롭아웃을 적용하지 않은 경우 훈련자료에 대한 오차는 에포크 100에서 0의 값으로 학습이 더 이상 진행되지 못하여 검증자료의 오차 또한 1.55%에서 멈추게 되는데 이는 드롭아웃을 사용한 경우의 최소오차인 1.24%보다 큰 값을 보였다. 반면 드롭아웃을 사용하는 경우 훈련자료에 대한 오차는 사용하지 않은 경우에 비하여 높았지만 꾸준히 감소하였으며 검증오차도 계속 감소하는 것을 알 수 있다.

Helmbold와 Long은 2017년에 드롭아웃이 기존의 일반화 방법인 가중치에 벌점을 부여하는 방법과는 다른 원리로 공적응 현상을 방지하는 근거를 밝혔으며 훈련자료 적합시에 지역최소 현상이 발생하지 않는다는 것을 보였다.

2.2. 하이브리드 드롭아웃

드롭아웃 알고리즘은 표본이 바뀔 때마다 새로운 모형을 사용하여 적합을 시도하고 손실(loss)을 계산한다. 이러한 방법은 공적응 현상을 막을 수는 있으나 표본마다 다른 모형을 적합하게 되어 Figure 3.1에서 살펴본 바와 같이 가중치들에 대한 추정치의 변동성은 드롭아웃을 사용하지 않는 경우보다 커지게 된다. 이러한 단점을 완화시키는 방법으로 하나의 모형에 하나의 표본을 적합하는 대신 둘 이상의 표본을 적합하도록 하여 추정치의 변동성을 줄이는 방법을 생각할 수 있다. 이를 통하여 가중치들의 추정치에

대한 변동성을 줄임과 동시에 모형의 안정성을 높혀 궁극적으로 수렴속도에 개선을 가져올 수 있으며 모의실험에서 사용한 자료들의 경우에는 검증자료에 대한 오차도 줄어드는 결과를 가져왔다.

하나의 모형에 적합하는 표본의 수를 k로 두면 $k = 1$인 경우가 드롭아웃이 되며 $k = 3$인 경우 하나의 모형에 3개의 표본을 적용하게 된다. 훈련은 미니배치 단위로 이루어지며 k가 2보다 큰 경우 주어진 각각의 모형에 미니배치에 배열된 순서에 따라 연속으로 k개의 표본을 학습에 사용하게 된다. Figure 3.3에는 $k = 3$, 즉, 하나의 모형에 3개의 표본을 적합시켜 얻어진 기울기(gradient)들의 평균을 사용하는 경우를 포함하고 있다. 앞 절에서 사용한 간단한 예제를 이용하여 이를 표현한 내용이 Figure 2.1에 포함되어 있다.

3. 모의실험

모의실험에서는 잘 알려진 두 종류의 이미지 자료인 MNIST자료와 CIFAR-10자료를 이용하여 기존의 드롭아웃 방법과 새롭게 제시한 하이브리드 드롭아웃 방법을 비교하였다. 드롭아웃은 다중퍼셉트론(multi-layer perceptron)구조의 완전연결신경망 뿐만 아니라 컨볼루셔널 신경망(convolutional neural network)과 long short-term memory (LSTM)이나 gated recurrent unit (GRU) 등의 순환신경망(recurrent neural network) 등에도 적용이 가능하다. 모의실험에 사용한 자료들이 모두 이미지 자료들 이므로 일반적인 완전연결신경망 보다는 컨볼루셔널 신경망이 더욱 적합하나 이들을 사용하는 경우에는 검증자료에 대한 오차가 매우 작아 방법간의 차이비교가 어려워지므로 오차가 상대적으로 큰 완전연결신경망을 이용하였다.

비교를 위하여 드롭아웃을 적용하지 않은 경우와 드롭아웃을 적용한 경우 그리고 드롭아웃 방법에서 각각의 모형이 훈련에 사용하는 표본의 수를 둘 이상으로 두는 경우 검증오차의 변화를 살펴보았다. 그리고 CIFAR-10자료를 이용하여 드롭아웃에서 입력값들과 각 노드의 선택확률 p와 각각의 모형이 훈련에 사용하는 표본의 수 k값의 여러 조합에 대한 검증오차를 통하여 조율 파라미터 p와 k의 영향을 살펴보았다.

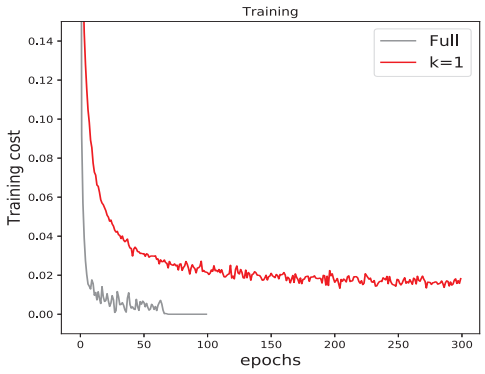
모든 분석은 구글에서 제공하는 Colab환경에서 텐서플로우(Tensorflow; https://tensorflow.org)패키지를 사용하여 수행하였다.

3.1. MNIST 자료

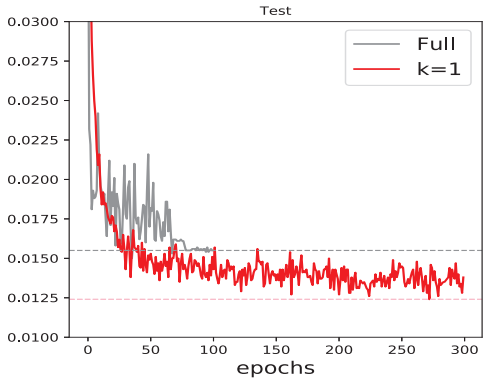
MNIST는 28×28 픽셀로 이루어진 흑백 이미지로 0부터 9까지의 숫자를 손으로 기록한 자료이며 학습을 위한 60,000개의 훈련자료와 10,000개의 검증자료로 이루어져 있다.

자료의 적합에는 은닉층이 2개 또는 3개이고 각각의 은닉층에 2,048개씩의 노드들을 포함하는 완전연결신경망 모형에 은닉층의 활성함수는 정류선형유닛을 사용하였다. 기타 조율 파라미터들의 값으로 학습률(learning rate)은 0.001로 두었고 초기값은 Xavier 초기값 (Glorot과 Bengio, 2010)을, 손실함수(loss function)는 교차-엔트로피(cross-entropy)를 사용하였다. 또한, 모든 드롭아웃 과정에서의 선택비율 p는 일반적인 경우 최적값으로 알려진 0.5로 설정하였으며 입력값과 은닉층의 노드에 모두 같은 값을 적용하였다.

앞의 2.1절에서 살펴본 바와 같이 MNIST자료에 대한 완전연결신경망모형에 드롭아웃을 적용하지 않는 경우 과적합의 발생과 함께 훈련자료의 오차가 에포크 100에서 0으로 수렴하면서 검증오차의 감소도 더 일어나지 않는 현상이 나타났다. 드롭아웃을 적용한 경우 훈련자료에 대한 오차가 급격하 감소하지 않는 대신 과적합 현상이 발생하지 않았으며 검증자료에 대한 최소오차도 드롭아웃을 적용하지 않는 경우

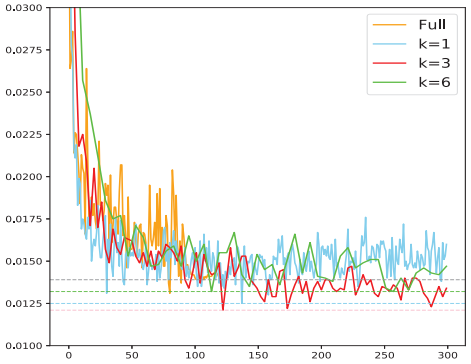


(a) Training cost for fully-connected network without dropout (Full) and fully-connected network with dropout ($k = 1$)

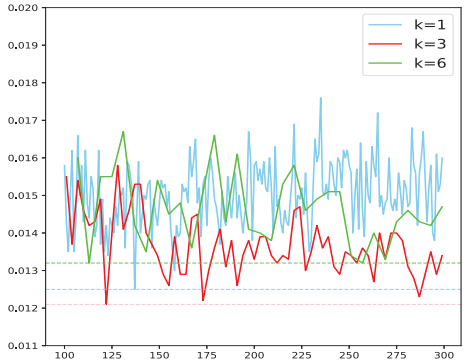


(b) Test error and minimum error (horizontal dotted line) for fully-connected network without dropout (Full) and fully-connected network with dropout ($k = 1$)

Figure 3.1. Dropout effect for MNIST with fully-connected network.



(a) Test error and minimum error (horizontal dotted line) of MNIST for fully-connected networks without dropout (full), with dropout ($k = 1$), hibrid dropout of $k = 3$ and $k = 6$



(b) Test error and minimum error (horizontal dotted line) of MNIST for fully-connected networks with dropout ($k = 1$), hybrid dropout of $k = 3$ and $k = 6$ between epoch 100 and 300

Figure 3.2. Test error of MNIST for $k = 1$ (dropout) and hybrid dropout with $k = 3$ and $k = 6$.

보다 작게 나타났다 (Figure 3.1).

은닉층이 3개이고 각각의 은닉층에 2,048개의 노드들로 구성된 모형의 검증오차와 관련된 결과가 Figure 3.2에 포함되어 있다. Figure 3.2(a)는 드롭아웃을 적용하지 않은 완전연결신경망 모형과 드롭아웃 및 하이브리드 드롭아웃 모형들의 검증오차 그래프이고 Figure 3.2(b)는 모형이 어느 정도 수렴한 이후인 에포크 100에서 300사이의 검증오차를 보여주고 있다. 수평축의 값은 $k = 1$인 경우에 해당되며 k값이 2이상인 경우 실제 에포크수는 수평축의 값을 k로 나눈 값이 된다. Table 3.1에는 에포크 100에서 300사이의 검증오차들의 최소값과 평균값, 그리고 표준편차값이 포함되어 있다. 그래프와 표의 결과에서 드롭아웃 방법이나 하이브리드 방법을 적용하는 경우 완전연결신경망에 비하여 검증오차가 더 작았

Table 3.1. Minimum, average, and standard deviation (STDEV) of test errors between epoch 100 and 300 for fully-connected network and models for $k = (1, 3, 6)$ with MNIST dataset

Network type	Minimum	Average	STDEV
Fully-connected	1.4%	-	-
$k = 1$	1.25%	1.53%	0.000808
$k = 3$	1.21%	1.35%	0.000517
$k = 6$	1.32%	1.45%	0.000713

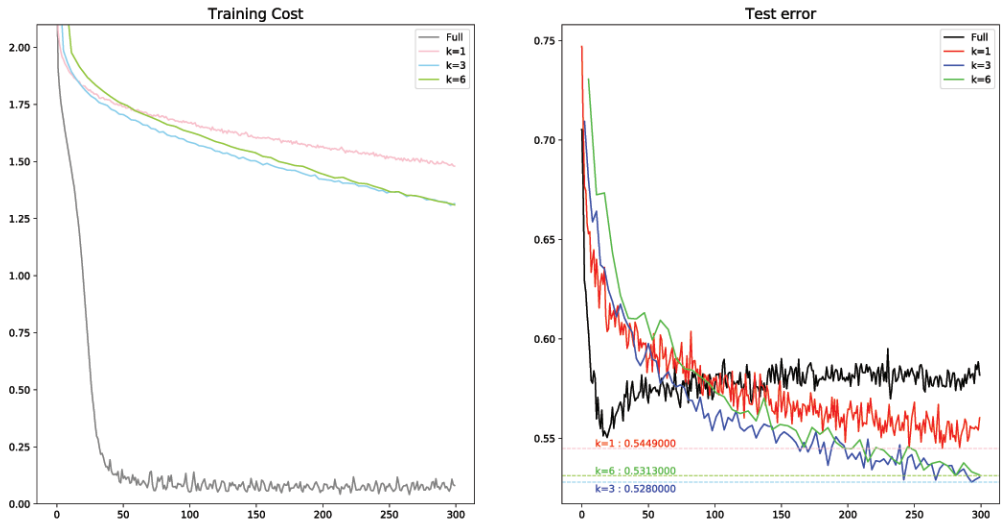


Figure 3.3. Training error (left plot) and Test error (right plot) of CIFAR-10 for fully-connected networks without dropout (full), with dropout ($k = 1$), hibrid dropout of $k = 3$ and $k = 6$.

으며 오차의 변동은 $k = 3$인 하이브리드 모형에서 가장 작았다. 특히 k값이 3인 경우 검증오차의 최소값과 평균, 그리고 변동이 가장 작아 배치의 크기 등에 따라 k 값을 적절히 조절하면 검증오차에 대한 최적의 결과를 얻을 수 있을 것으로 판단된다.

3.2. CIFAR-10 자료

두 번째로 고려한 CIFAR-10 자료는 10개의 서로 다른 객체들로 구성되어 있는 $32 \times 32 \times 3$ 차원의 색상 이미지 자료로 훈련을 위한 50,000개의 훈련자료와 10,000개의 검증자료로 이루어져 있다. 색상 이미지 자료는 RGB (red, green, blue) 3개의 채널(자료의 세 번째 차원)로 구성되어 있으며 사용된 모형의 입력에 적합하도록 Luma 코딩 방법(luma $= 0.2126 \times R + 0.7153 \times G + 0.0722 \times B$)을 적용하여 회색조(grayscale)로 변환한 후 벡터화하여 사용하였다. 완전연결신경망 모형의 은닉층의 수는 3개인 경우만을 고려하였다. 은닉층의 활성함수와 그 이외의 조율 파라미터들의 값은 MNIST 자료에서 사용한 것과 동일한 값들을 사용하였다. 드롭아웃 및 하이브리드 모형에서의 노드 선택확률 p도 같은 값인 0.5를 그대로 적용하였다.

Figure 3.3의 왼쪽에는 모형들의 훈련자료에 대한 오차들의 그래프가 포함되어 있다. 기본적으로 모형의 검증오차가 앞의 MNIST와 달리 매우 크기 때문에 완전연결신경망모형과 기타 드롭아웃 방법들 간의 훈련오차에 큰 차이가 나타났다. 기본적인 드롭아웃 방법에 비하여 하이브리드 방법의 훈련오차가

Table 3.2. Minimum, average, and standard deviation (STDEV) of test errors between epoch 100 and 300 for fully-connected network and models for $k = (1, 3, 6)$ with CIFAR-10 dataset

Network type	Minimum	Average	STDEV
Fully-connected	55.05%	-	-
$k = 1$	54.49%	56.27%	0.008658
$k = 3$	52.38%	54.44%	0.008369
$k = 6$	53.13%	54.21%	0.011770

Table 3.3. Test errors for combinations of $p = (0.3, 0.5, 0.7)$ and $k = (\text{Full}, 1, 3, 10, 20)$

Accept probability	Fully-connected	$k = 1$	$k = 3$	$k = 10$	$k = 20$
$p = 0.3$		64.43%	60.17%	60.65%	62.34%
$p = 0.5$	55.04%	54.49%	52.38%	53.18%	56.51%
$p = 0.7$		52.86%	52.79%	52.64%	55.19%

모든 구간에서 작게 나타났으나 $k = 3$인 경우 에포크 50 이상에서 대부분의 경우 제일 작았으며 에포크가 진행될수록 $k = 6$인 경우도 $k = 3$인 경우와 비슷한 값으로 수렴하는 결과를 보였다. 같은 그림 오른쪽에 포함된 검증오차들의 그래프에서도 드롭아웃이나 하이브리드 방법들에서는 훈련자료에 대한 오차의 그래프와 비슷한 양상을 보였으며 $k = 3$인 경우 오차의 크기가 가장 작으면서 동시에 변동도 일반적인 드롭아웃 및 다른 k값들의 경우보다 작았다. 검증오차의 최소값은 $k = 3$인 경우 52.8% 였으나 기존의 드롭아웃 방법의 최소값 54.49%에 비하여 감소폭이 크지 않아 평균적인 검증오차의 크기와 관계없이 효과는 한정되게 나타나는 것으로 보인다. Table 3.2에는 앞의 MNIST 자료에서와 같이 에포크 100에서 300사이의 검증오차들의 최소값과 평균값, 그리고 표준편차값이 포함되어 있다. 이 자료에서도 드롭아웃 방법이나 하이브리드 방법을 적용하는 경우 완전연결신경망에 비하여 검증오차가 더 작았으며 오차의 변동 또한 $k = 3$인 하이브리드 모형에서 가장 작았으나 상대적인 크기는 크지 않았다.

은닉층의 수는 그대로 두고 각각의 은닉층에 포함된 노드의 수를 2,048개에서 2배로 늘린 4,096개인 모형도 고려하였으나 검증자료에 대한 오차가 평균적으로 약 2% 정도 감소하는 것 이외에 다른 특성들에 큰 차이가 없어 포함하지 않았다.

3.3. p와 k에 대한 시뮬레이션: CIFAR-10 자료

MNIST와 CIFAR-10자료에 대한 결과들을 통하여 하나의 모형이 훈련과정에서 고려하는 표본의 수 k가 둘 이상인 경우 검증오차가 줄어드는 효과가 있었으나 단조증가하지 않는 것으로 나타났다. 본 절에서는 CIFAR-10자료를 이용하여 검증오차에 미치는 p값과 k값의 효과를 알아보기 위하여 $p = (0.3, 0.5, 0.7)$의 세가지 경우와 $k = (1, 3, 10, 20)$의 네가지 경우의 조합들에 대하여 간단한 시뮬레이션을 수행하였다. 사용된 신경망 구조와 기타 조율 파라미터들의 값은 앞 절의 CIFAR-10의 예제와 동일하게 두었다.

Figure 3.4에는 p와 k의 조합들에 해당하는 예측오차와 드롭아웃 방법들을 적용하지 않은 완전연결신경망모형의 예측오차의 플롯을 포함하고 있다. 기존의 드롭아웃 방법이나 하이브리드 방법을 적용한 모든 경우 과적합현상은 발생하지 않았으나 $p = 0.3$인 경우 모든 k 값에서 검증오차가 완전연결모형의 오차값보다 크게 나타났으며 평균적으로 모형의 크기가 완전연결모형의 70% 정도가 되는 $p = 0.7$인 경우 다른 경우와 비교하여 수렴속도가 매우 빨라 드롭아웃을 사용하지 않은 경우와 비슷하게 국소최소값(local minima)으로 수렴할 위험이 커질 수 있는 것으로 판단된다. 추가적으로 $p = 0.7$인 경우 $k = 20$인 경우를 제외하고 모든 경우에서 검증오차의 최소값은 53%를 넘지 않았으나 오차의 변동이 크

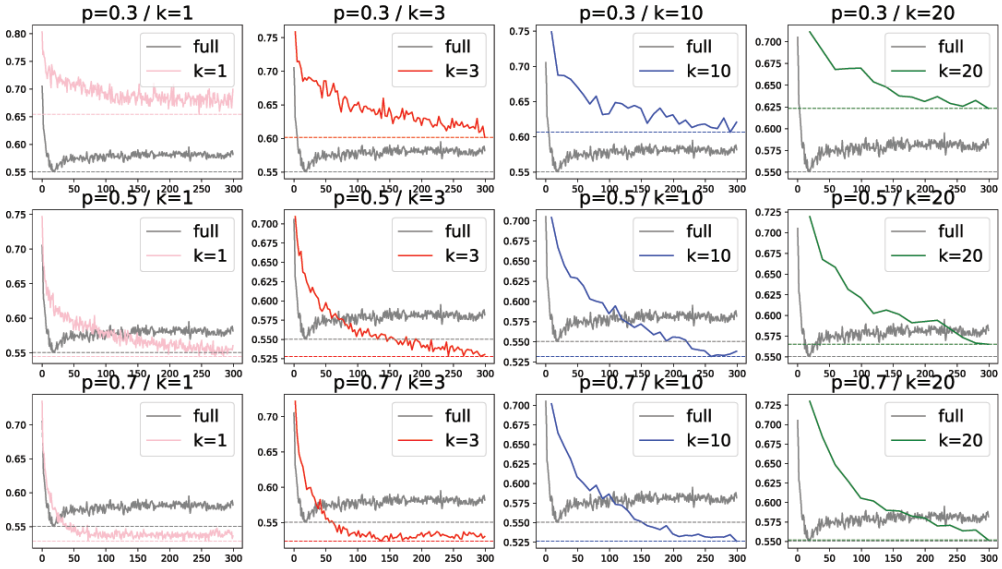


Figure 3.4. Plots of test errors for combinations of $p = (0.3, 0.5, 0.7)$ and $k = (1, 3, 10, 20)$.

고 안정적으로 수렴하지 못하는 현상을 관찰할 수 있었으며 오차값도 $p = 0.5$인 경우와 비교하여 상대적으로 큰 값들이 자주 나타났다. Table 3.3에서 볼 수 있듯이 $p = 0.5$이고 $k = 3$인 경우 예측오차의 최소값이 가장 작았으며 같은 k값에서 p가 어느 정도 큰 값을 가져도 예측오차에 큰 변화는 없었다. 결과적으로 두 파리미터의 값에 따라 모형의 수렴 특성과 최소 예측오차가 영향을 받는 것으로 나타났다.

4. 결론

본 논문에서는 심층신경망모형에서 발생할 수 있는 과적합 현상을 방지하는 드롭아웃 방법을 사용하는 경우 모형의 모수 추정치들의 변동성이 크고 수렴과정에서 변동이 줄어들지 못하는 단점을 해결하는 방안으로 훈련과정에서 각각의 모형 적합에 둘 이상의 표본을 사용하는 방법을 제안하고 이를 하이브리드 드롭아웃이라고 하였다.

잘 알려진 MNIST와 CIFAR-10 이미지 자료를 일반적인 완전신경망모형을 사용하여 새로운 방법과 기존의 드롭아웃 방법을 적용한 결과 새로운 방법을 적용하면 기존의 드롭아웃 방법의 단점인 추정치의 변동성과 수렴과정에서 변동이 감소하지 못하는 단점을 완화시키고 더 나아가 검증자료에 대한 예측력도 높아지는 것을 확인하였다. CIFAR-10자료를 이용하여 각각의 모형 적합에 사용되는 표본수 k와 입력 및 은닉층의 노드 선택확률 p의 조합이 모형의 적합 특히 예측오차에 미치는 영향을 살펴보았다. 우선 k의 경우 모든 선택확률에서 예측오차가 k값에 따라 단조증가하지 않아 격자 탐색(grid search) 등을 통하여 적절할 값을 선택할 필요가 있음을 알았다. p값의 경우 전체모형과의 크기 차이가 작은 경우(시뮬레이션에서는 $p = 0.7$인 경우) 드롭아웃을 사용하지 않은 경우과 같이 훈련자료에 대한 오차가 급격히 줄어들어 과적합 현상이 나타나는 것을 알 수 있었다. p와 k의 조합이 미치는 영향에 대한 상세한 정보를 얻으려면 추가적인 모의실험이 필요하다고 하겠다.

기존의 드롭아웃 방법과 새롭게 제시한 하이브리드 드롭아웃 방법 모두 각 에포크에서의 검증오차(훈련오차도 같은 현상을 보임)의 변동폭이 미니배치 경사하강법에 비하여 심하게 나타남을 알 수 있다. 이는

하나 또는 k개의 표본마다 적합되는 모형이 동일하지 않아 가중치의 추정치 변동이 심하여 발생하는 것으로 짐작할 수 있으며 이러한 현상을 줄여 모형의 예측오차 변동성을 줄이는 방법으로 가중치의 추정치를 구하는 과정에서 미니배치안의 표본들에 대한 평균을 계산하기 전에 실제 모형의 p값을 각 표본의 가중치에 곱한 값을 사용하는 것을 생각할 수 있을 것이다. 완전연결신경망 이외에 이미지 자료에 적합한 컨볼루셔널 신경망 등을 적용하는 경우 다른 결과가 나타날 수도 있어 이에 대한 추가적인 연구도 필요하다.

References

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks, Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

Helmbold, D. and Long, P. (2017). Surprising properties of dropout in deep networks, *Proceedings of Machine Learning Research*, **65**, 1–24.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580.

Ioffe, S. and Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *Proceeding ICML'15 Proceedings of the 32nd International Conference on International Conference on Machine Learning*, **37**, 448–456.

Nowlan, S. and Hinton, G. (1992). Simplifying neural networks by soft weight-sharing, *Neural Computation*, **4**, 473–493.

Salakhutdinov, R. and Mnih, A. (2008). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo, *In Proceedings of the 25th International Conference on Machine Learning*, ACM.

Srivastava, N. (2013). Improving Neural Networks with Dropout. Master's thesis, University of Toronto.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, **15**, 1929–1958.

Wan, L., Zeiler, M., Zhang, S., Cun, Y. and Fergus, R. (2013). Regularization of neural networks using DropConnect, *Proceedings of the 30th International Conference on Machine Learning*, **28**, 1058–1066.

Xiong, H., Barash, Y. and Frey, B. (2011). Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context, *Bioinformatics*, **27**, 2554–2562.

하이브리드 드롭아웃

박종선[a,1] · 이명규[a]

[a]성균관대학교 통계학과

(2019년 10월 2일 접수, 2019년 11월 1일 수정, 2019년 11월 5일 채택)

요 약

수 많은 모수들을 가지고 있는 방대한 심층신경망은 매우 강력한 기계학습 방법이지만 모형의 과도한 융통성으로 인하여 과적합문제를 내포하고 있다. 드롭아웃 방법은 크기가 큰 신경망의 과적합 문제를 해결하는 다양한 방법들 중 하나이며 매우 효과적인 방법으로 알려져 있다. 드롭아웃 방법은 훈련과정에서 각각의 표본에 다른 모형을 적용하는데 이들 모형은 입력과 은닉층의 노드들을 무작위로 제거한 모형들 중에 임의로 선택된다. 본 연구에서는 임의로 선택된 모형에 둘 이상의 표본을 적용하여 모형의 가중치들에 대한 추정치의 안정성을 높이는 하이브리드 드롭아웃 방법을 제시하였다. 실제 자료를 이용한 시뮬레이션 결과 노드의 선택확률과 모형의 적합에 사용되는 표본의 수를 적절하게 선택하여 기존의 방법에 비하여 추정치의 변동성이 감소시킬 수 있었으며 동시에 검증자료에 대한 최저오차도 줄일 수 있음을 보였다.

주요용어: 심층신경망, 드롭아웃, 공적응, 과적합, 하이브리드방법

[1]교신저자: (03063) 서울특별시 종로구 성균관로 25-2, 성균관대학교 통계학과. E-mail: cspark@skku.edu