기계 학습 연구 저널 15 (2014) 1929-1958 제출일 11/13; 6월 14일 게시됨

드롭아웃: 신경망의 과적합을 방지하는 간단한 방법

니티시 스리바스타바 제프리 힌튼 알렉스 크리제프스키

NITISH@CS.TORONTO.EDU HINTON@CS.TORONTO.EDU KRIZ@CS.TORONTO.EDU

일리야 수츠케버 ILYA@CS.TORONTO.EDU

Ruslan Salakhutdinov 토론토 컴퓨터 과학 대학의학과

10 Kings College Road, Rm 3302 토론토, 온타리오, M5S 3G4, 캐나다.

에디터: 요슈아 벤지오

RSALAKHU@CS.TORONTO.EDU

추상적인

매개변수가 많은 심층 신경망은 매우 강력한 기계 학습 시스템입니다. 그러나 과적합은 이러한 네트워크에서 심각한 문제입니다. 또한 대규모 네트워크는 사용 속도가 느리기 때문에 테스트 시간에 다양한 대규모 신경망의 예측을 결합하여 과적합을 처리하기 어렵습니다. 드롭아웃은 이 문제를 해결하기 위한 기술입니다. 핵심 아이디어는 훈련 중에 신경망에서 단위(연결과 함께)를 무작위로 떨어뜨리는 것입니다. 이렇게 하면 유닛이 너무 많이 공동 적응하는 것을 방지할 수 있습니다. 교육 중에 기하급수적으로 다양한 "얇아진" 네트워크에서 샘플을 드롭아웃합니다. 테스트 시간에 더 작은 가중치를 가진 단일 비얇은 네트워크를 사용하여 이러한 모든 얇은 네트워크의 예측을 평균화하는 효과를 쉽게 근사화할 수 있습니다. 이는 과적합을 크게 줄이고 다른 정규화 방법에 비해 크게 개선됩니다. 우리는 드롭아웃이 시각, 음성 인식, 문서 분류 및 전산 생물학의 지도 학습 작업에서 신경망의 성능을 향상시켜 많은 벤치마크 데이터 세트에서 최신 결과를 얻음을 보여줍니다.

키워드: 신경망, 정규화, 모델 조합, 딥러닝

1. 소개

심층 신경망에는 여러 개의 비선형 은닉층이 포함되어 있어 입력과 출력 간의 매우 복잡한 관계를 학습할 수 있는 매우 표현력이 풍부한 모델이 됩니다. 그러나 훈련 데이터가 제한되어 있으면 이러한 복잡한 관계 중 많은 부분이 샘플링 노이즈의 결과이므로 훈련 세트에는 존재하지만 동일한 분포에서 추출된 경우에도 실제 테스트 데이터에는 존재하지 않습니다. 이는 과적합으로 이어지며 이를 줄이기 위한 많은 방법이 개발되었습니다. 여기에는 검증 세트의 성능이 악화되기 시작하는 즉시 교육을 중단하고 L1 및 L2 정규화 및 소프트 가중치 공유와 같은 다양한 종류의 가중치 패널티를 도입하는 것이 포함됩니다(Nowlan 및 Hinton, 1992).

무제한 계산을 통해 고정 크기 모델을 "정규화"하는 가장 좋은 방법은 가능한 모든 매개 변수 설정의 예측을 평균화하고 각 설정에 다음과 같이 가중치를 부여하는 것입니다.

⃝c 2014 Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever 및 Ruslan Salakhutdinov.

SRIVASTAVA, HINTON, KRIZHEVSKY, SUTSKEVER 및 SALAKHUTDINOV

그림 1: 드롭아웃 신경망 모델. 왼쪽: 2개의 은닉층이 있는 표준 신경망. 오른쪽: 왼쪽 네트워크에 드롭아웃을 적용하여 생성된 가늘어진 네트워크의 예입니다. 교차 유닛이 삭제되었습니다.

Srivastava, Hinton, Krizhevsky, Sutskever and Salakhutdinov

패턴, 대칭, 예술이(가) 표시된 사진

자동 생성된 설명



훈련 데이터가 주어진 사후 확률. 이것은 때때로 단순하거나 작은 모델에 대해 아주 잘 근사화될 수 있지만(Xiong et al., 2011; Salakhutdinov and Mnih, 2008), 훨씬 적은 계산을 사용하여 베이지안 금 표준의 성능에 접근하고 싶습니다. 우리는 매개변수를 공유하는 학습된 모델의 기하급수적인 수의 예측에 대해 균등하게 가중된 기하 평균을 근사함으로써 이를 수행할 것을 제안합니다.

모델 조합은 거의 항상 기계 학습 방법의 성능을 향상시킵니다. 그러나 대규모 신경망의 경우 개별적으로 훈련된 많은 신경망의 출력을 평균화한다는 명백한 아이디어는 엄청나게 비용이 많이 듭니다. 여러 모델을 결합하는 것은 개별 모델이 서로 다를 때 가장 유용하며 신경망 모델을 다르게 만들기 위해서는 서로 다른 아키텍처를 갖거나 다른 데이터에 대해 훈련되어야 합니다. 각 아키텍처에 대한 최적의 하이퍼파라미터를 찾는 것은 어려운 작업이고 각각의 대규모 네트워크를 교육하려면 많은 계산이 필요하기 때문에 다양한 아키텍처를 교육하는 것은 어렵습니다. 또한 대규모 네트워크에는 일반적으로 많은 양의 교육 데이터가 필요하며 데이터의 서로 다른 하위 집합에서 서로 다른 네트워크를 교육하는 데 사용할 수 있는 데이터가 충분하지 않을 수 있습니다. 많은 다른 대규모 네트워크를 훈련할 수 있었다고 해도 테스트 시간에 모두 사용하는 것은 신속하게 응답하는 것이 중요한 애플리케이션에서는 실행 불가능합니다.

드롭아웃은 이 두 가지 문제를 모두 해결하는 기술입니다. 과적합을 방지하고 기하급수적으로 다양한 신경망 아키텍처를 효율적으로 결합하는 방법을 제공합니다. "드롭아웃"이라는 용어는 신경망에서 유닛(숨겨진 것과 보이는 것)을 떨어뜨리는 것을 말합니다. 장치를 삭제한다는 것은 그림 1과 같이 들어오고 나가는 모든 연결과 함께 장치를 네트워크에서 일시적으로 제거하는 것을 의미합니다. 삭제할 장치의 선택은 무작위입니다. 가장 간단한 경우 각 단위는 다른 단위와 독립적인 고정 확률 p로 유지되며 여기서 p는 유효성 검사 세트를 사용하여 선택하거나 간단히 0.5로 설정할 수 있습니다. 작업. 그러나 입력 단위의 경우 최적의 유지 확률은 일반적으로 0.5보다 1에 더 가깝습니다.

## present

* + 1. At training time (b) At test time

라인, 텍스트, 폰트, 도표이(가) 표시된 사진

자동 생성된 설명

그림 2: 왼쪽: 확률 p로 존재하고 가중치 w로 다음 계층의 단위에 연결된 훈련 시간의 단위. 오른쪽: 테스트 시간에 단위는 항상 존재하며 가중치에 p를 곱합니다. 테스트 시간의 출력은 훈련 시간의 예상 출력과 동일합니다.

신경망에 드롭아웃을 적용하는 것은 신경망에서 "가늘어진" 네트워크를 샘플링하는 것과 같습니다. 얇은 네트워크는 드롭아웃에서 살아남은 모든 장치로 구성됩니다(그림 1b). ㅏ

n 단위의 신경망은 2n개의 얇은 신경망의 모음으로 볼 수 있습니다. 이러한 네트워크는 모두 가중치를 공유하므로 매개변수의 총 개수는 여전히 O(n2)입니다. 또는

더 적은. 각 교육 사례의 각 프레젠테이션에 대해 새로운 얇은 네트워크가 샘플링되고 교육됩니다. 따라서 드롭아웃이 있는 신경망을 교육하는 것은 광범위한 가중치 공유가 있는 2n개의 얇은 네트워크 모음을 교육하는 것으로 볼 수 있습니다. 여기서 각 얇은 네트워크는 교육을 받습니다.

아주 드물게, 만약 있다면.

테스트 시간에 기하급수적으로 많은 얇아진 모델의 예측을 명시적으로 평균화하는 것은 불가능합니다. 그러나 매우 간단한 대략적인 평균화 방법이 실제로 잘 작동합니다. 아이디어는 테스트 시간에 드롭아웃 없이 단일 신경망을 사용하는 것입니다. 이 네트워크의 가중치는 훈련된 가중치의 축소 버전입니다. 단위가 훈련 중에 확률 p로 유지되면 해당 단위의 나가는 가중치는 그림 2와 같이 테스트 시간에 p로 곱해집니다.

훈련 시간에 단위를 떨어뜨리는 데 사용되는 분포)는 테스트 시간의 실제 출력과 동일합니다. 이 스케일링을 수행하면 공유 가중치가 있는 2n 네트워크를 테스트 시간에 사용할 단일 신경망으로 결합할 수 있습니다. 우리는 네트워크를 훈련시키는 것을 발견했습니다.

탈락을 방지하고 테스트 시간에 이 근사 평균화 방법을 사용하면 다른 정규화 방법을 사용한 훈련에 비해 다양한 분류 문제에서 일반화 오류가 상당히 낮아집니다.

드롭아웃의 개념은 피드 포워드 신경망에만 국한되지 않습니다. Boltzmann Machines와 같은 그래픽 모델에 보다 일반적으로 적용될 수 있습니다. 이 논문에서는 dropout Restricted Boltzmann Machine 모델을 소개하고 이를 표준 Restricted Boltzmann Machines(RBM)과 비교합니다. 우리의 실험은 드롭아웃 RBM이 특정 측면에서 표준 RBM보다 우수함을 보여줍니다.

본 논문은 다음과 같이 구성되어 있다. 섹션 2는 이 아이디어의 동기를 설명합니다. 섹션 3은 관련 이전 작업을 설명합니다. 섹션 4는 드롭아웃 모델을 공식적으로 설명합니다. 섹션 5는 드롭아웃 네트워크 훈련을 위한 알고리즘을 제공합니다. 섹션 6에서는 다른 도메인의 문제에 드롭아웃을 적용하고 다른 형태의 정규화 및 모델 조합과 비교하는 실험 결과를 제시합니다. 섹션 7에서는 신경망의 다양한 속성에 대한 드롭아웃의 영향을 분석하고 드롭아웃이 네트워크의 하이퍼파라미터와 상호 작용하는 방법을 설명합니다. 섹션 8은 드롭아웃 RBM 모델을 설명합니다. 섹션 9에서 우리는 탈락을 소외시키는 아이디어를 탐구합니다. 부록 A에서는 실용적인 가이드를 제공합니다.