

Deep Sparse Rectifier Neural Networks

≡ AI 키워드	
📅 날짜	@2024년 9월 23일
≡ 콘텐츠	논문
≡ 태그	유런



- 2011년, Xavier Glorot
- ReLU가 소개된 논문
 - $ReLU(x) = \max(0, x)$
- sparsity와 모델 학습의 관계



- dying ReLU 문제
 - Leaky ReLU 등의 대안 존재

Abstract

sigmoid 혹은 hyperbolic tangent network(tanh)에 대한 대안
sparse 데이터, 준지도학습에 적합

Introduction

computational neuroscience 모델과 머신러닝 모델이 activation의 관점에서 연결될 수 있음

rectifier activation function: $\max(0, x)$

신경망이 3개 층 이상일 때 효과적

실험

이미지 분류 태스크에서 logistic sigmoid와 비교

denoising autoencoder pre-training 진행 후 tanh와 비교

결과

비지도 사전학습이 없을 때 최고 성능 달성

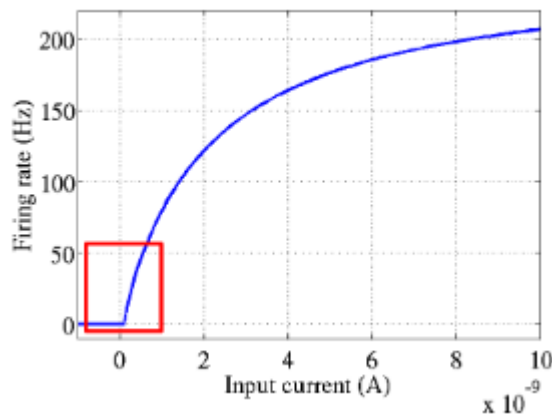
비라벨링 데이터 많을 때 준지도 학습을 도울 수 있음

희소 네트워크를 생성하므로 생물학적 뉴런의 응답에 가까

Background

sigmoid, tanh

relu는 실제 뉴런의 작동 방식(=정보를 sparse하게 처리)을 기반으로 설계되었음



Advantages of Sparsity

1. 정보 분리: dense data는 서로 얽혀있어 분리하기 어려움
2. 효율적이고 다양한 크기의 표현
3. 선형 분리성
4. distributed but sparse

Deep Rectifier Networks

Rectifier Neurons

$$\text{rectifier}(x) = \max(0, x)$$

Advantages

쉽게 희소 표현을 얻음

각각 이산적으로는 선형적 → gradient vanishing이 나타나지 않음(sigmoid나 tanh의 비선형성에서 발생하는)

계산량 적음

Potential Problems

0이 많아지면 역전파를 방해할 수 있다고 가정

$\text{softplus}(x) = \log(1 + e^x)$ 와 비교 → 가정 맞지 않음

gradient가 일정 path를 따라 전파되는 한 괜찮음

L1과 같은 regularization을 같이 쓴다면 추가적인 희소성이 발생

2배 더 많은 은닉층 필요

Unsupervised Pre-training

Experimental Study

Image Recognition

데이터셋: MNIST, CIFAR10, NISTP, NORB

Table 1: **Test error on networks of depth 3.** Bold results represent statistical equivalence between similar experiments, with and without pre-training, under the null hypothesis of the pairwise test with $p = 0.05$.

Neuron	MNIST	CIFAR10	NISTP	NORB
<i>With</i> unsupervised pre-training				
Rectifier	1.20%	49.96%	32.86%	16.46%
Tanh	1.16%	50.79%	35.89%	17.66%
Softplus	1.17%	49.52%	33.27%	19.19%
<i>Without</i> unsupervised pre-training				
Rectifier	1.43%	50.86%	32.64%	16.40%
Tanh	1.57%	52.62%	36.46%	19.29%
Softplus	1.77%	53.20%	35.48%	17.68%

- unsupervised pre-training 없을 때는 rectifier가 가장 좋은 결과
- 근데 편차가 그리 크지 않음

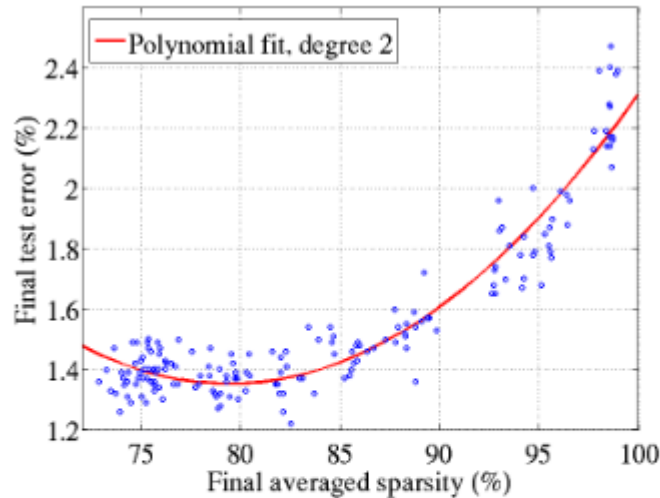


Figure 3: **Influence of final sparsity on accuracy.** 200 randomly initialized deep rectifier networks were trained on MNIST with various L_1 penalties (from 0 to 0.01) to obtain different sparsity levels. Results show that enforcing sparsity of the activation does not hurt final performance until around 85% of true zeros.

- 70%에서 85% 사이의 sparsity를 가질 때 가장 robust

Sentiment Analysis

데이터셋: 레스토랑 리뷰

Table 2: Test RMSE and sparsity level obtained by 10-fold cross-validation on OpenTable data.

Network	RMSE	Sparsity
No hidden layer	0.885 ± 0.006	$99.4\% \pm 0.0$
Rectifier (1-layer)	0.807 ± 0.004	$28.9\% \pm 0.2$
Rectifier (3-layers)	0.746 ± 0.004	$53.9\% \pm 0.7$
Tanh (3-layers)	0.774 ± 0.008	$00.0\% \pm 0.0$

- rectifier 적용한 것이 가장 좋은 결과

Conclusion

- 깊은 신경망에서 뉴런은 대부분 선형 영역에서 작동하며, 희소성과 잘 결합할 수 있음.
- 비지도 사전학습을 실행한 것과 사전학습을 실행하지 않은 것의 차이를 줄여줌 → 학습 과정에서 더 적합한 지점을 찾을 수 있음
- 50%에서 80%의 희소성을 가지는 네트워크를 얻을 수 있음(뇌가 가질 것 같다고 가정하는 95%에서 99%의 희소성과 관련 있음)
- 텍스트 기반 작업, 특히 감성 분석과 같은 높은 희소성을 가진 데이터에 적합하다는 점에서 유망함