Feature Learning and Mapping using Deep Learning Approaches Hamid Mohammadi CSLU Seminar 2015-02-03

Outline

- Artificial Neural Networks (ANNs)
- MNIST classification task
- Unsupervised Feature Learning
 - various ANN architectures
 - Evaluations
- Supervised Feature Mapping
 - various ANN architectures
 - Evaluations

ANN Architecture

- ANN is composed of multiple layers
- Layers perform non-linear transformations



http://bengio.abracadoudou.com/lectures/old/tex_ann.pdf

Backpropagation



Backpropagation

• Criterion for ANN

- Mean Squared Error:
 - Error=(ŷ-y)^2
- Cross-entropy

Backpropagation

- Regularization:
 - L1: Criterion = CE + |W|
 - L2: Criterion = $CE + W^2$
 - Dropout: randomly omitting subsets of features at each iteration with probability p=[0.0, 0.5]

MNIST Corpus

- 28x28 pixels, pixel values range from 0 to 1
- Contains 70,000 images
 - 50,000 training set
 - 10,000 validation set
 - 10,000 test set



• Task: Classify 10 digit classes

ANN for MNIST

• Two-layer ANN with 100 hidden units





Feature Learning

• Usual Machine Learning applications have two steps



Feature Learning

• Usual Machine Learning applications have two steps



Feature Learning



ANN for Feature Learning

- How to use ANNs for Feature Learning?
- ONE solution is to use networks called Autoencoders
- These networks try to reconstruct the input
- They are unsupervised, no labels needed



- These networks try to reconstruct the input
- The model's equation:
 x_rec=g(W'g(Wx+b)+b')
- where W' =transpose (W)
 (tied weights)



- Implementation:
 - Python 2.7
 - Theano 0.6

• AE-1000 weights (1000 hidden units)

				.0	
	6.63		2.14 A.4-		2 dece
			1 1 1 T		
100	1. A.	55.0	1 1.5 %	2	* . ·
1. A.				in the second	
27 - 1				4.5	
		1.10			
	1.	r _e i	S.		
		5.			
	1.4		1		
	1994 - 1994 1994 - 1994 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1994 - 1		7 2		23 (S)

• AE-1000-L1 weights

	12 A. 19 197.00	1 X X 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	· · · · · · · · · · · · · · · · · · ·	23,2012
经金融 化化学	(1) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
		Contra a la contra	10 1 5 1 1 1	10000
	S. C. D. P. S. M. C.	and have	STORES STORES	61.664
All and a starting	24 Y W Startes	A COLORADO	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	D
· · · · · · · · · · · · · · · · · · ·	A Statistics	Level Port	and a start of the	11000
and the second second	A 2 2 4 33 12	1. 白水 白水	1. 人名法尔尔	Sec. 2. 14
		A CONTRACT CONTRACTOR		1 20 20 20 20
a martine and	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1.1 4221	1.312 3.4	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1
115 Mar 19 20 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1. 15 1.	1	2.00	1. 1. 20
And the Property Sector	A start and	Alt and a second		1.1
Sectors of Stational Labor.	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Distance in the second	COLOR STOCK	1.500
	Sela Cost Cost	Part and a second	14275年4月1日	
	(1) 分子上的目前的。	E. Carlos Marco	7.57 . 30 . 2 . 5 . 6	20. 100-10
·公共的 ···································	1. ····································	1.1.1	1877 A. B. Barley	1. 1. 2. 1. 1
NAME OF GROOM	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		1. 新闻的发展	
1 m 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Status - 12		Prove Prove Sta	12.00
アイション ない ない		122 124.19	21.32 MA. 55	and se
the second second		CALCULATION OF MALE		12.00
and the state of the	3 3 4 4 4 W 18 6	14 S 1 7 7	1. 1. 1. 1. 1. A.	The States
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	and the second second	Real Provest	and the state of the	Cont.
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Sector State	1. 1. 1. 1. 1. 1. TAV	2.40.01.25.2	
ALC: NOT STREET	1.	A. 1	State of the Store Add	A X 19 C
的关系的 。在1995年1月19日	Sel LANNER	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	1. A. B. A. B.	12.0
A PARTY PARTY A		1 States And States	10 5 mil 10 10	
Sec. 27 4. 28 42	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	新闻的 和1995	1. 1. 1. 1. 1. 1. 1.	151.273
在以前的行行。	7 8 10 Car 15	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		2 Call
1. 行政 方法法法法	Server Server	(13) ····································	12.36日 出现的	
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	A-201 248	Carl Lands	1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	1. AL
	and the second second second	The second second	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1.000
	Carl State	Part of the second	12/11/2012	SY2. 18.
13.1	State 1 185	1.5. C	A. A. A. A. A.	1. 2.6
A CONTRACTOR	STATISTICS.	1000	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1.9 5 . 4
Total Andrew Party of the	March 199		1 http://www.com	
State of the state of the state	1. 1. 1. 1. 1. 1.	No. Carlo Garde	Section de sal	19.5.17
and the second second	100 20 000	Sec. 324 184 18	新活动 · 200	1. 19. 19.10
19 4 19 19 19 19 19 19 19 19 19 19 19 19 19	the second second	A State State	A CAN BUSH	1.00

• AE-1000-L2 weights

n de la competencia de la comp	01:00-1				
			•		
			$\sum_{i=1}^{L}$		
		10			15

• AE-1000-Dropout-0.1

								•	
						1.			
という		-				•	1.11		•
•	Hires &		1		•		-		•
	•			ans.					- 1
•	T .			5					
V .				•		•		1. 1. 1	$\hat{\boldsymbol{\lambda}}$
N.	C.			•					
	1		T N	1			15.15 15.45		
			and a second				1		



Pretraining ANN

• AE training



Reconstructed Input

- No pre-training: Initializing ANN weights from random numbers
- Pre-training: Initializing ANN weights from AE weights



ANN for MNIST

• ANN with 1000 hidden units



• ANN pre-trained with regular AE



• ANN pre-trained with regular AE-L1-0.001



• ANN pre-trained with regular AE-L2-0.001



• ANN pre-trained with regular AE-Dropout-0.1



• ANN pre-trained with regular AE-Dropout-0.5



Stacked Autoencoders

Different Levels of Abstraction





3rd layer "Objects"

2nd layer "Object parts"

1st layer "Edges"

Pixels

https://deeplearningworkshopnips2010.files.wordpress.com/2010/09/nips10-workshop-tutorial-final.pdf

Evaluation

Model	Notes	Reconstruction	Classification Error %
ANN 1000	_	_	1.84%
ANN 1000	Pre-trained AE	0.00101	1.65%
ANN 1000	Pre-trained AE-L1-0.001	0.00131	1.72%
ANN 1000	Pre-trained AE-L20.001	0.00194	1.66%
ANN 1000	Pre-trained AE-Dropout-0.1	0.00256	1.66%
ANN 1000	Pre-trained AE-Dropout-0.5	0.01556	1.70%
DNN 1000-1000	_	-	1.74%
DNN 1000-1000	AE-Dropout-0.1-0.2	0.00411	1.46%
DNN 1000-1000-1000	_	-	1.72%
DNN 1000-1000-1000	AE-Dropout-0.1-0.2-0.3	0.00952	1.40%
100 supervised tra lr=0.1, bate	aining iterations, 15 un h=10, SGD	nsupervised trainin lr=0.05, batch=10	g iterations, , SGD

Evaluation

- 49,000 images for unsupervised training
- 1000 image for supervised training

Model	Notes	Classification Error %		
ANN 100	_	12.92%		
ANN 1000	_	13.47%		
DNN 1000-1000-1000	_	12.71%		
ANN 1000	Pre-trained AE-Dropout-0.1	10.36%		
DNN 1000-1000-1000	AE-Dropout-0.1-0.2-0.3	8.75%		
100 supervised training iterations,15 unsupervised training iterations,lr=0.1, batch=10, SGDlr=0.05, batch=10, SGD				

- The usual application of DNN is classification
- We will propose some training methods for mapping (regression)
- These have the potential to be used in voice conversion

- In this study, we will perform our evaluations on MNIST
- x: half left images
- y: half right images

11543 75353 55906 35200

• Simple ANN mapping



• The previously proposed approach for VC



- We propose Correlated AEs
- Background
- Canonical Correlation Analysis
- Two random variables X1 and X2
- find W1 and W2 such that W1.X and W2.X2 are maximally correlated

- Deep CCA
- The hidden layer outputs have the highest correlation Canonical Correlation Analysis



- DCCA does not have good reconstruction
- We propose join the cost of reconstruction and DCCA Canonical Correlation Analysis



• 1000 labeled, 49,000 unlabeled

Model	Notes	Reconstruction
ANN 100	_	0.0724
ANN 100	AE-Dropout-0.1	0.0554
ANN 100	AE-Dropout-0.1+DCCA	0.0463

400 supervised training iterations, lr=0.1, batch=20, SGD 15 unsupervised training iterations, lr=0.1, batch=20, SGD