Fastai Study Group

Cleveland AI Group (CAIG)
October 22, 2018 - Week 4

Event Hosts

Michael Kudlaty
@Michael Kudlaty

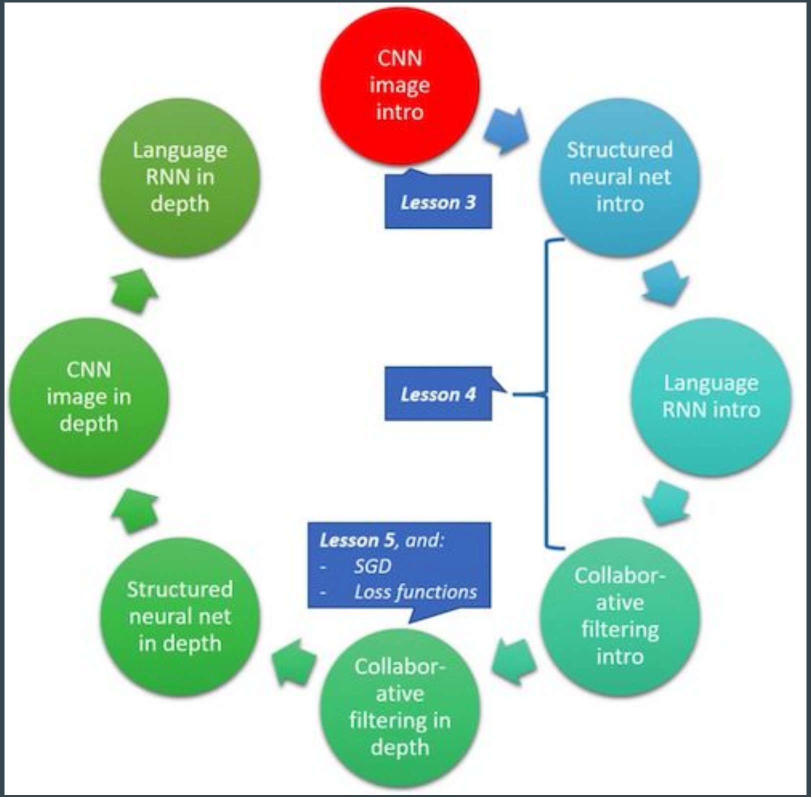
Jason Mancuso
@jvmancuso

Michał Wojczulis
@Michał

Brendan Mulcahy
@Brendan Mulcahy

Agenda

- Lesson 4 key concepts
  - Dropout
  - Structured Data
  - Embedding Matrices
  - Natural Language Processing
  - Intro to Collaborative Filtering
- Discussion/questions
- Presentations (if any)



Adapted from:
https://medium.com/@hiromi_suenaga/deep-learning-2-part-1-lesson-4-2048a26d58aa

Dropout

“For each mini-batch, we throw away a different random half of activations in that layer. It forces it to not overfit.”

Conv2

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0		0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0
0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1		1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	2	3	3	5	4	2		0	0	0	0	0
0	0	0	0	0	1	1	2	1	1		1	1	2	3	3	5	7	4	2	0	0	0	0	0
0	0	2	2	2	2	3	5	5	4	4	3	3	2	2		4	7	7	4	1	0	0	0	0
1	2	5	5		4	4	5	5	4	3	3	3	3	2	2	2	5	8	6	2	0	0	0	0
2	2	3	4	3	2	2	2	2	1	1	1	1	1	2	2	1		7	8	4	0	0	0	0
0	0	1	0	0	0	0		0	0	0	0	0	0	0	1	0	1	5	9	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	4	9	7	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0		0	1	1	0	3	8		3	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	3	2	2	3	7	8	4	0	0	0
0	0	0	0		0	0	0	1	1	1	2	3	4	5	5	3	3	4	7	8	3	0	0	0
0	0	0	0	0	0	1	1	3	4	5	6	5	5	5	4	2	2	4	8	7	2	0	0	0
0	0	0	0	0	1	4	5	6	6	5	4	3	2	2	1	0	2	5	9	6	1	0	0	0
0	0	0	0	0	2	5	6	4	3	2	1	1	0	0	0	1	2	6	8	5	0	0	0	0
0	0	0	1	1	2	3	3	1	0	0	0	0	0	0	0	2	3	8	7	3	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	3	5	8	6	1	0	0	0	0

Maxpool

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	3	3	0	0
0	0	1	2	1	1	2	3	7	4	0
2	5	4	5	5	4	3	2	7	8	2
2	4	3	2	2	1	1	2	1	9	5
0	0	0	0	0	0	0	1	1	9	7
0	0	0	0	1	2	4	5	3	7	8
0	0	1	5	6	6	5	5	2	9	7
0	1	2	6	4	2	1	0	3	8	5
0	0	0	1	0	0	0	2	7	8	1

Dropout (cont.)

- Typically about 50% chance to drop each activation:
  - Flip a coin for each activation, if it comes up tails, replace with zero
  - Double all the remaining activations to keep the average the same
- Not a strong intuition/understanding of dropout (i.e. not mathematically rigorous)
  - If your validation loss is higher than your training loss (i.e. overfitting),
    then increase dropout (e.g. try 70%)
  - If your validation loss is lower than your training loss (i.e. underfitting),
    Then decrease dropout (e.g. try 25%)
- Dropout is not used during inference time (i.e. after training, we don't continue to perform dropouts)
- Forces the network to spread its understanding out across several weights and not rely very heavily on particular activations

Structured Data

	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
0	1	5	2015-07-31	5263	555	1	1	0	1
1	2	5	2015-07-31	6064	625	1	1	0	1
2	3	5	2015-07-31	8314	821	1	1	0	1
3	4	5	2015-07-31	13995	1498	1	1	0	1
4	5	5	2015-07-31	4822	559	1	1	0	1

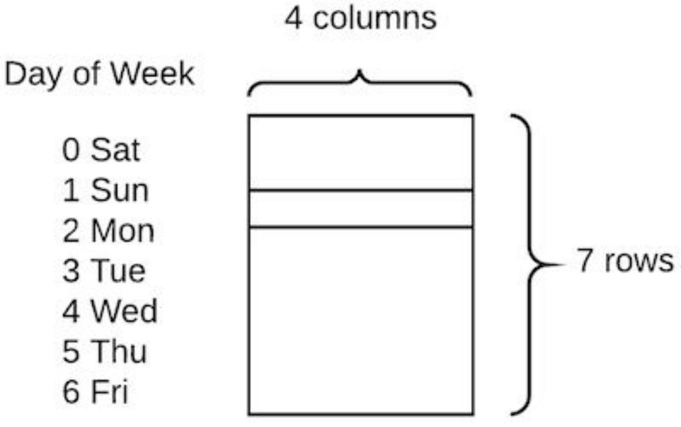
	Store	StoreType	Assortment	CompetitionDistance	CompetitionOpenSinceMonth	CompetitionOpenSinceYear	Promo2	Promo2SinceWeek	Promo2SinceYear
0	1	c	a	1270.0	9.0	2008.0	0	NaN	NaN
1	2	a	a	570.0	11.0	2007.0	1	13.0	2010.0
2	3	a	a	14130.0	12.0	2006.0	1	14.0	2011.0
3	4	c	c	620.0	9.0	2009.0	0	NaN	NaN
4	5	a	a	29910.0	4.0	2015.0	0	NaN	NaN

Data Types

- Categorical - has categories/levels/groups
  - Store Type
  - Item Assortment
- Continuous - floating point numbers
  - CompetitionDistance - represents distance to the nearest competitor
  - MeanTemperature
- Year/DayOfWeek
  - Can be treated as categorical or continuous
  - Often works better to treat them as categorical so that the network can treat each item separately
- Cardinality - the number of possible levels in a category (e.g. the cardinality of DayOfWeek is 8, because there is 7 days + 1 for Unknown)

Embedding Matrices

- As we have already seen, it makes sense to use continuous values in a network
- How do we use categorical values?
    - We create an embedding matrix which has 1 row per level, and a (somewhat) arbitrary number of columns (Jeremy chooses about half the cardinality, so 4 columns for DayOfWeek)
    - Each column contains a random initialized float
    - For each row in our input data set, we replace the entry with the corresponding row from the embedding matrix
    - Then perform gradient descent on the weights from that embedding row
    - Basically, we create a few floats per category and use those in the network



Natural Language Processing (NLP)

- Old area of research but is a relatively new area for deep learning
- Given some input text:
  - Produce more output text
  - Generate a sentiment analysis

sample_model(m, "<CAT> csni <SUMM> algorithms that")

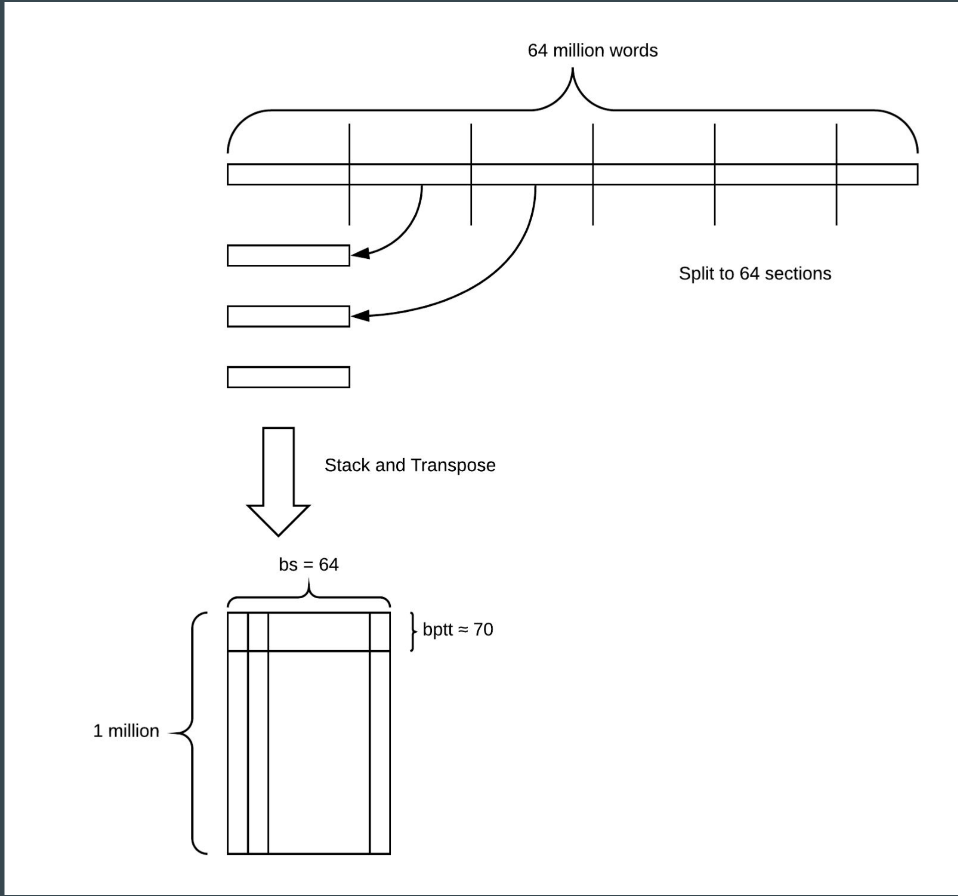
...use the same network as a single node are not able to achieve the same performance as the traditional network - based routing algorithms . in this paper , we propose a novel routing scheme for routing protocols in wireless networks . the proposed scheme is based ...

NLP (cont.)

- Create an embedding matrix for every word (token) in the sample data
- Tokenization - how/where to split text into tokens
  - Typically, split on whitespace and punctuation (for English)
  - Examples:
    - “This is a sample sentence!” becomes [“this”, “is”, “a”, “sample”, “sentence”, “!”]
    - “wasn’t” becomes [“was”, “n’t”]
  - Jeremy uses a tokenizer called spacy_tok, but there are many tokenization algorithms

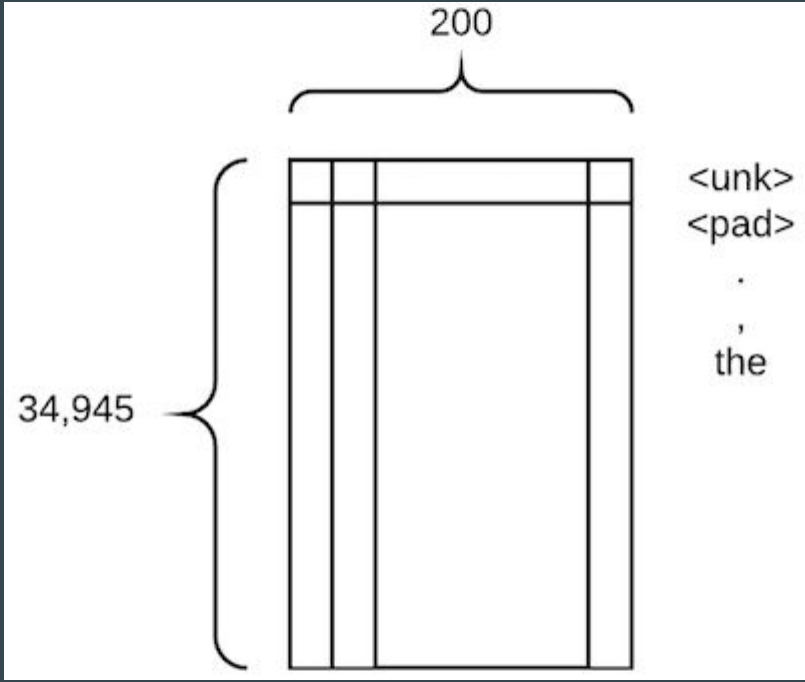
NLP (cont.)

- Connect all text samples together
- bs = batch size = # of tokens in a batch/section
- bptt = back prop through time = # number of batches to use at 1 time



NLP Embedding Matrix & Architecture

- Rows = # of unique tokens
- Cols = ~200, more than DayOfWeek because we need ability to store the nuance
- Architecture = Recurrent Neural Network (RNN), different from what we have seen before, CNN
- RNN are to time as CNN are to space, where time in this case is a word sequence



Collaborative Filtering

- Given a bunch of users and their ratings of movies, can we predict ratings of movies that a particular user has not yet seen? i.e. Netflix

	userId	movieId	rating	timestamp
0	1	31	2.5	1260759144
1	1	1029	3.0	1260759179
2	1	1061	3.0	1260759182
3	1	1129	2.0	1260759185
4	1	1172	4.0	1260759205

movieId	1	110	260	296	318	356	480	527	589	593	608	1196	1198	1270	2571
userId															
15	2.0	3.0	5.0	5.0	2.0	1.0	3.0	4.0	4.0	5.0	5.0	5.0	4.0	5.0	5.0
30	4.0	5.0	4.0	5.0	5.0	5.0	4.0	5.0	4.0	4.0	5.0	4.0	5.0	5.0	3.0
73	5.0	4.0	4.5	5.0	5.0	5.0	4.0	5.0	3.0	4.5	4.0	5.0	5.0	5.0	4.5
212	3.0	5.0	4.0	4.0	4.5	4.0	3.0	5.0	3.0	4.0	NaN	NaN	3.0	3.0	5.0
213	3.0	2.5	5.0	NaN	NaN	2.0	5.0	NaN	4.0	2.5	2.0	5.0	3.0	3.0	4.0
294	4.0	3.0	4.0	NaN	3.0	4.0	4.0	4.0	3.0	NaN	NaN	4.0	4.5	4.0	4.5
311	3.0	3.0	4.0	3.0	4.5	5.0	4.5	5.0	4.5	2.0	4.0	3.0	4.5	4.5	4.0
380	4.0	5.0	4.0	5.0	4.0	5.0	4.0	NaN	4.0	5.0	4.0	4.0	NaN	3.0	5.0
452	3.5	4.0	4.0	5.0	5.0	4.0	5.0	4.0	4.0	5.0	5.0	4.0	4.0	4.0	2.0
468	4.0	3.0	3.5	3.5	3.5	3.0	2.5	NaN	NaN	3.0	4.0	3.0	3.5	3.0	3.0
509	3.0	5.0	5.0	5.0	4.0	4.0	3.0	5.0	2.0	4.0	4.5	5.0	5.0	3.0	4.5
547	3.5	NaN	NaN	5.0	5.0	2.0	3.0	5.0	NaN	5.0	5.0	2.5	2.0	3.5	3.5
564	4.0	1.0	2.0	5.0	NaN	3.0	5.0	4.0	5.0	5.0	5.0	5.0	5.0	3.0	3.0
580	4.0	4.5	4.0	4.5	4.0	3.5	3.0	4.0	4.5	4.0	4.5	4.0	3.5	3.0	4.5
624	5.0	NaN	5.0	5.0	NaN	3.0	3.0	NaN	3.0	5.0	4.0	5.0	5.0	5.0	2.0

Resources

- Wiki lesson 4 https://forums.fast.ai/t/wiki-lesson-4/9402
- Lesson 4 notes Link
- Course page http://course.fast.ai/
- Course forums http://forums.fast.ai/
- Cleveland Tech Slack https://cleveland-tech.herokuapp.com/
    - Join #deep_learning channel
    - Ask questions or share articles
- AI Saturdays https://nurture.ai/ai-saturdays
- AI Saturdays guide Link
- AI Saturdays forums https://ai6forums.nurture.ai/
- CAIG Website https://clevelandaigroup.github.io/

Thank You to Our Sponsors!

TECH ELEVATOR

python SOFTWARE FOUNDATION

Questions?