Deep Learning for NLP Live Training

Efficient Processing of Natural Language with Artificial Neural Networks

The Pomodoro Technique

-- Francesco Cirillo, 1980s

Today's Training Structure:

1. three pomodoros (25 minutes + 5 minute break)
2. bonus five-minute break
3. three further pomodoros

Questions typically handled at break end

POLL

Where are you?

- The Americas
- Europe / Middle East / Africa
- Asia-Pacific
- Extra-Terrestrial Space

POLL

What are you?

- Developer / Engineer
- Scientist / Analyst / Statistician / Mathematician
- Combination of the Above
- Other

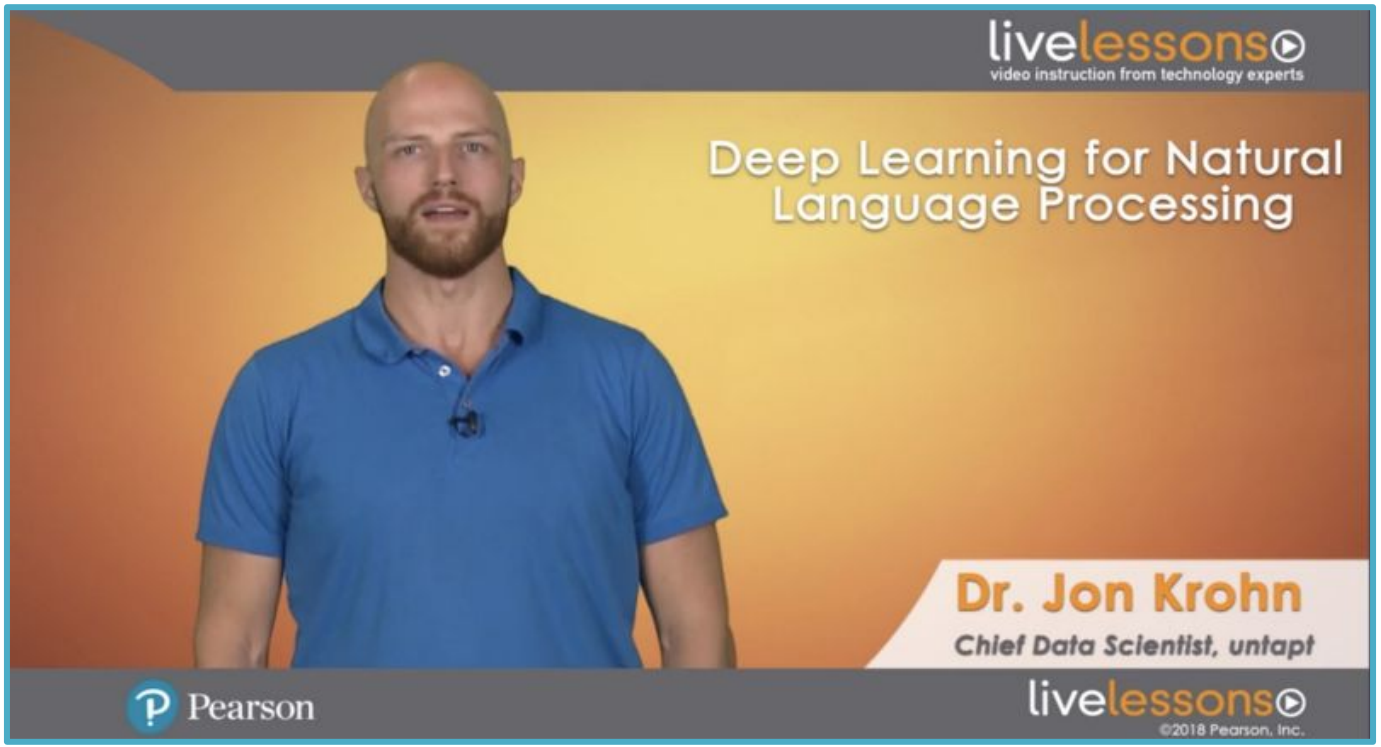
Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

Deep Learning for NLP

1. **The Power and Elegance of Deep Learning for NLP**
   - Intro to Deep Learning for NLP
   - Review of Deep Learning Theory
   - Vector-Space Embeddings
   - Creating Word Vectors with word2vec
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

Dig Deeper (“*reference NLP LiveLessons*”)



Search *“Deep Learning for Natural Language Processing”* or *“Jon Krohn”* in Safari

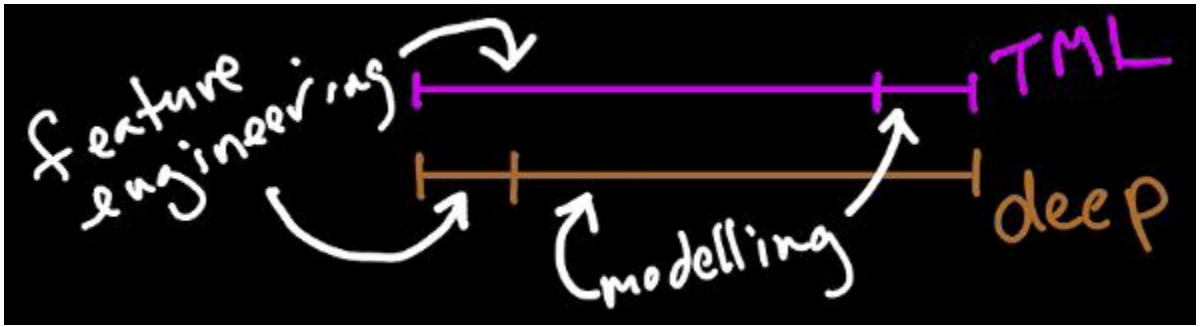
Deep Learning for NLP

1. **The Power and Elegance of Deep Learning for NLP**
   - **Intro to Deep Learning for NLP** *(reference NLP LiveLessons sections 1.1 to 1.3)*
   - Review of Deep Learning Theory
   - Vector-Space Embeddings
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

Two Core Concepts

1. Deep Learning
2. Natural Language Processing (NLP)

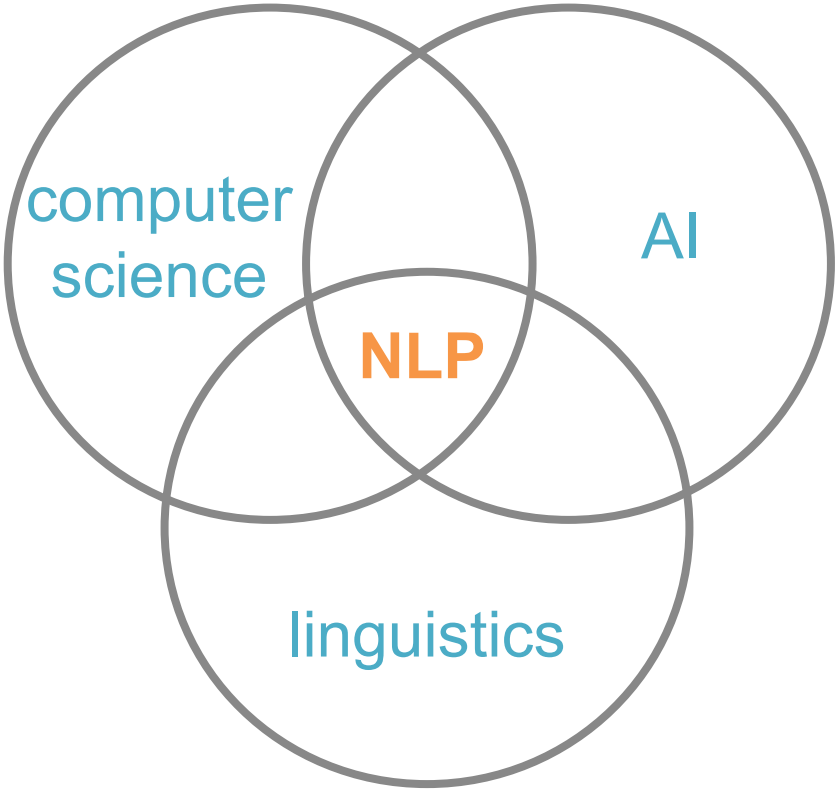
Traditional ML vs Deep Learning



Two Core Concepts

1. Deep Learning
2. Natural Language Processing (NLP)

Natural Language Processing



NLP Applications

Easy

- spell checking
- synonym suggestions
- keyword search

NLP Applications

Intermediate

- reading level
- extracting information
- predicting next words
- classification
- sequence generation
- time series analysis

NLP Applications

Complex

- machine translation
- question-answering
- chatbots

POLL

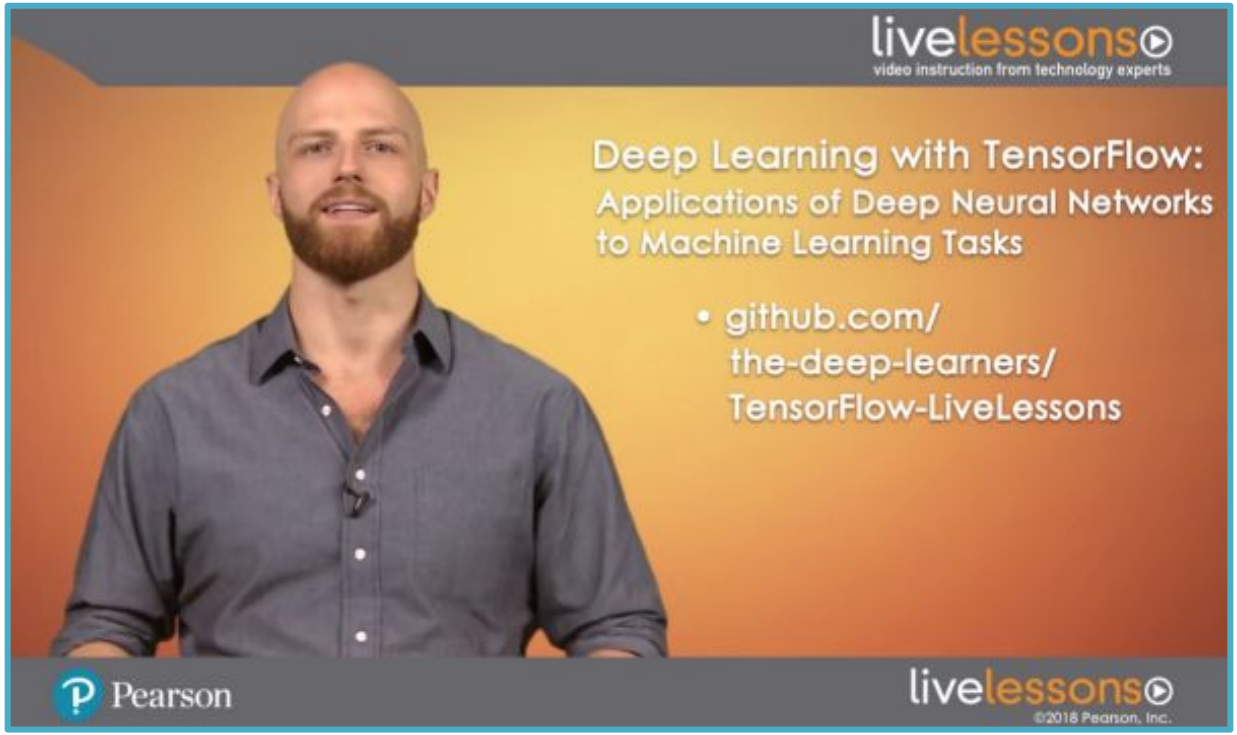
Which of the following is unrelated to Natural Language Processing?

- AI
- Computer Science
- Linguistics
- Neuro-Linguistic Programming

Deep Learning for NLP

1. **The Power and Elegance of Deep Learning for NLP**
   - Intro to Deep Learning for NLP
   - **Review of Deep Learning Theory** *(reference NLP LiveLessons section 1.5, OR Deep Learning with TensorFlow LiveLessons)*
   - Vector-Space Embeddings
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

Deep Learning Fundamentals (assumed)



Search *"Deep Learning with TensorFlow"* or *"Jon Krohn"* in Safari

Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

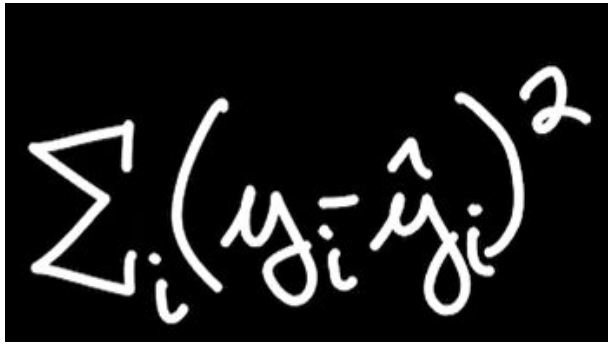
Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost



Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost
- cross-entropy

Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost
- cross-entropy

***Stochastic* Gradient Descent**

- mini-batch size
- learning rate
- second-order, e.g., Adam

Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost
- cross-entropy

***Stochastic* Gradient Descent**

- mini-batch size
- learning rate
- second-order, e.g., Adam

Backpropagation

Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost
- cross-entropy

***Stochastic* Gradient Descent**

- mini-batch size
- learning rate
- second-order, e.g., Adam

Backpropagation

Initialization

- Glorot normal
- Glorot uniform

Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost
- cross-entropy

***Stochastic* Gradient Descent**

- mini-batch size
- learning rate
- second-order, e.g., Adam

Backpropagation

Initialization

- Glorot normal
- Glorot uniform

Layers

- dense / fully-connected
- convolutional
- max-pooling
- flatten
- softmax

Your Arsenal

Neurons

- sigmoid
- tanh
- ReLU

Cost Functions

- quadratic cost
- cross-entropy

***Stochastic* Gradient Descent**

- mini-batch size
- learning rate
- second-order, e.g., Adam

Backpropagation

Initialization

- Glorot normal
- Glorot uniform

Layers

- dense / fully-connected
- convolutional
- max-pooling
- flatten
- softmax

Avoiding Overfitting

- L1/L2 Regularization
- dropout
- data expansion

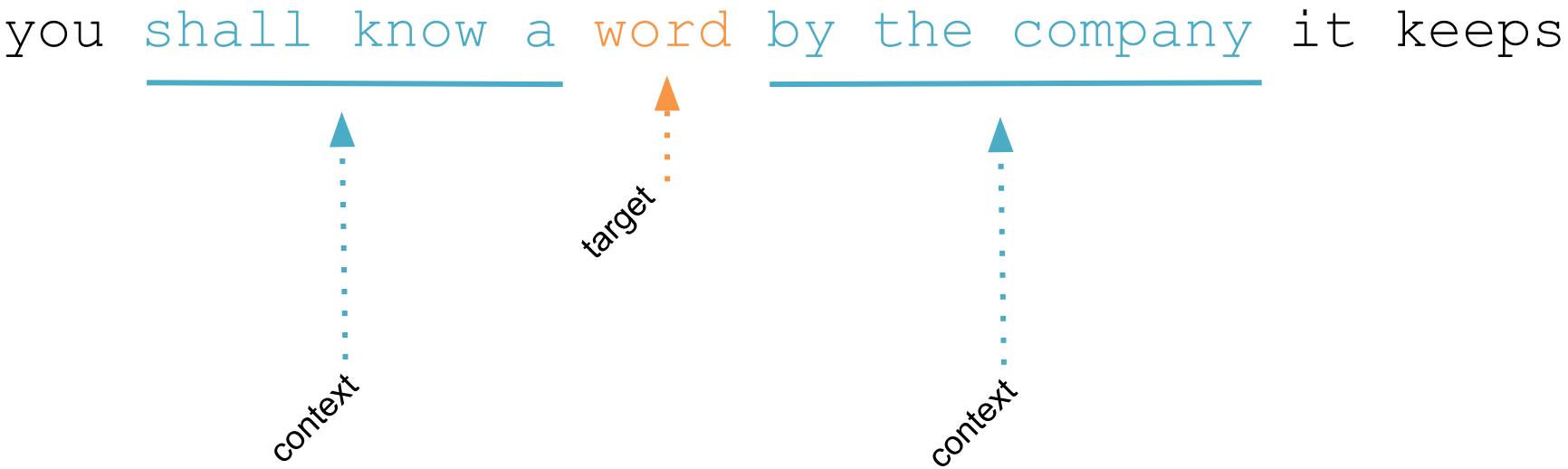
Deep Learning for NLP

1. **The Power and Elegance of Deep Learning for NLP**
   - Intro to Deep Learning for NLP
   - Review of Deep Learning Theory
   - **Vector-Space Embeddings** *(reference NLP LiveLessons sections 2.1 and 2.2)*
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

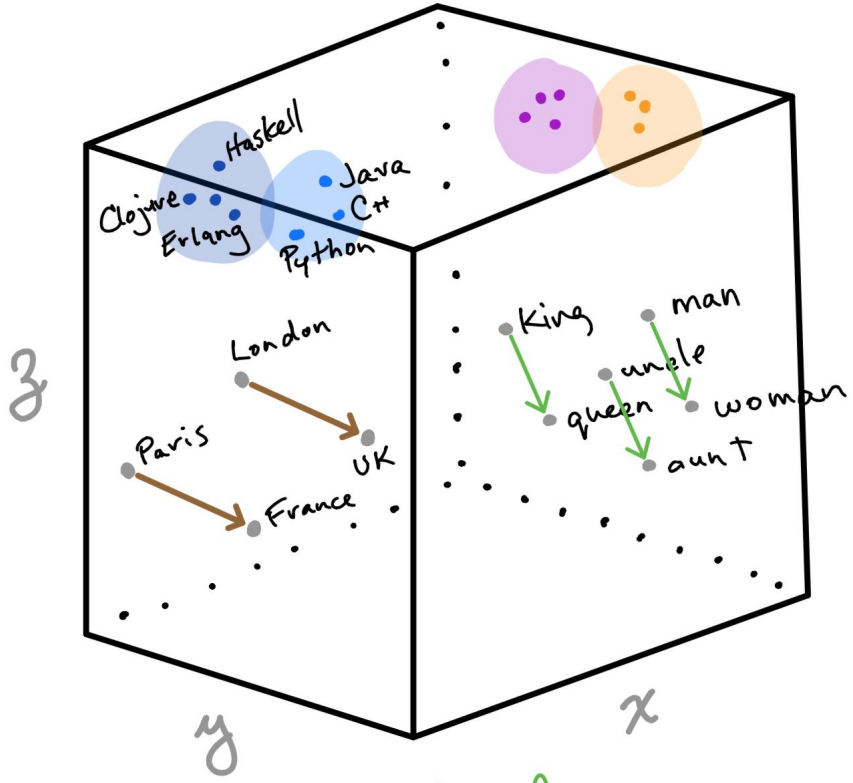
One-Hot Word Representations

word	The	cat	sat	on	the	mat.
the	1	0	0	0	1	0
cat	0	1	0	0	0	0
on	0	0	0	1	0	0
⋮						
N_{unique_words}						

Vector Representations of Words



Vector Representations of Words



n-dimensional space

Word Vector Arithmetic

$$v_{king} - v_{man} + v_{woman} = v_{queen}$$

$$v_{bezos} - v_{amazon} + v_{tesla} = v_{musk}$$

$$v_{windows} - v_{microsoft} + v_{google} = v_{android}$$

$$v_{cu} - v_{copper} + v_{gold} = v_{au}$$

Word Representations

One-Hot	Vector-Based
lack nuance	extremely **nuanced**
handle new words poorly	seamlessly incorporate **new words**
subjective	**driven by** natural language **data**
laborious, manual taxonomies	fully-**automatic**
word similarity ignored	**word similarity** = closeness in space
unwieldy with large vocabulary	accommodate **large vocabularies**

Vector-Space Embeddings

[word2viz demo]

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. **Modeling Natural Language Data**
   - Preprocessing Natural Language for word2vec
   - Document Classification with a Dense Net
   - Document Classification with a ConvNet
3. Recurrent and Advanced Neural Networks

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. **Modeling Natural Language Data**
   - **Preprocessing Natural Language for word2vec** *(reference NLP LiveLessons sections 2.4 and 3.1)*
   - Document Classification with a Dense Net
   - Document Classification with a ConvNet
3. Recurrent and Advanced Neural Networks

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. **Modeling Natural Language Data**
   - Preprocessing Natural Language for word2vec
   - **Document Classification with a Dense Net** *(reference NLP LiveLessons section 3.3)*
   - Document Classification with a ConvNet
3. Recurrent and Advanced Neural Networks

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. **Modeling Natural Language Data**
   - Preprocessing Natural Language for word2vec
   - Document Classification with a Dense Net
   - **Document Classification with a ConvNet** *(reference NLP LiveLessons section 3.4)*
3. Recurrent and Advanced Neural Networks

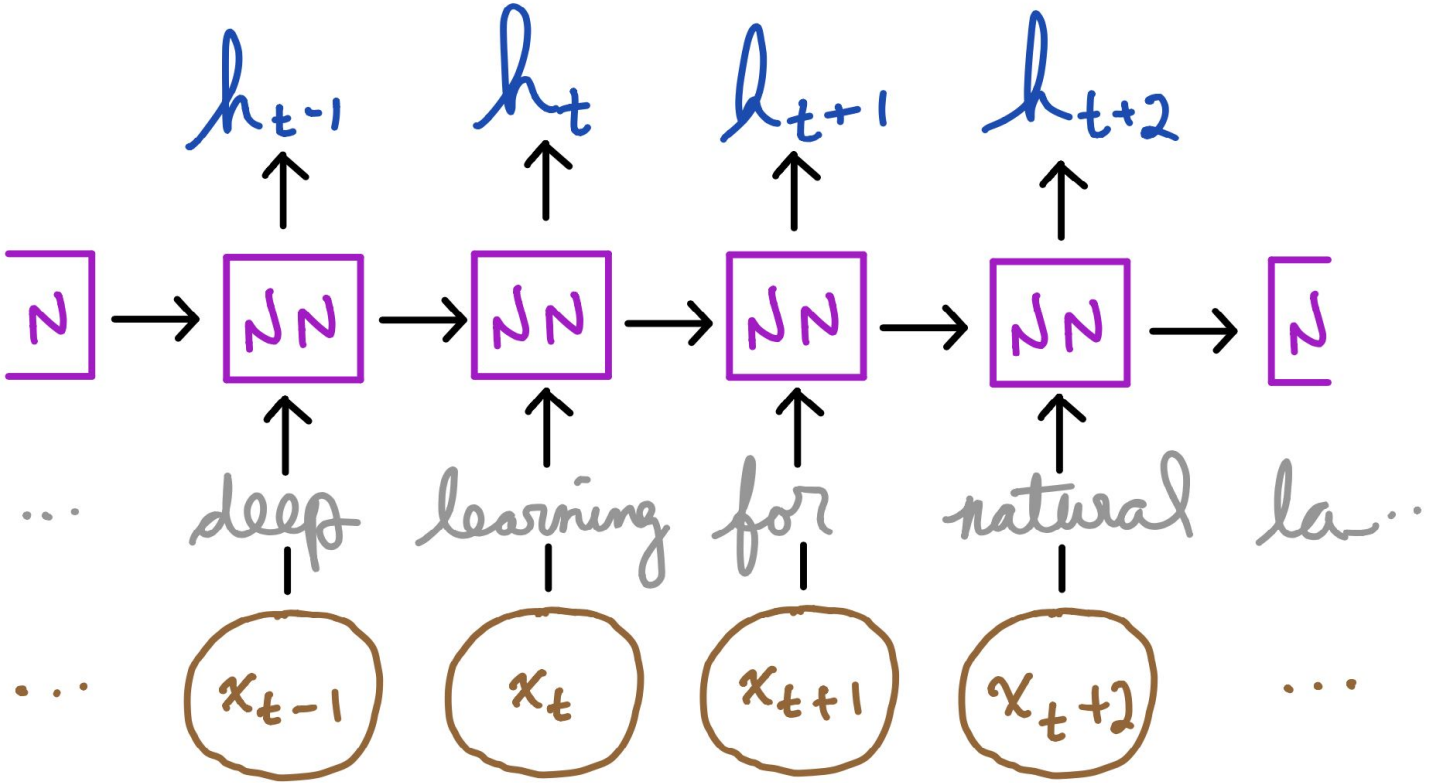
Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
   - Recurrent Neural Networks
   - LSTMs
   - Bi-Directional LSTMs
   - Stacked LSTMs
   - Parallel Network Architectures

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
   - Recurrent Neural Networks *(reference NLP LiveLessons 4.1 and 4.2)*
   - LSTMs
   - Bi-Directional LSTMs
   - Stacked LSTMs
   - Parallel Network Architectures

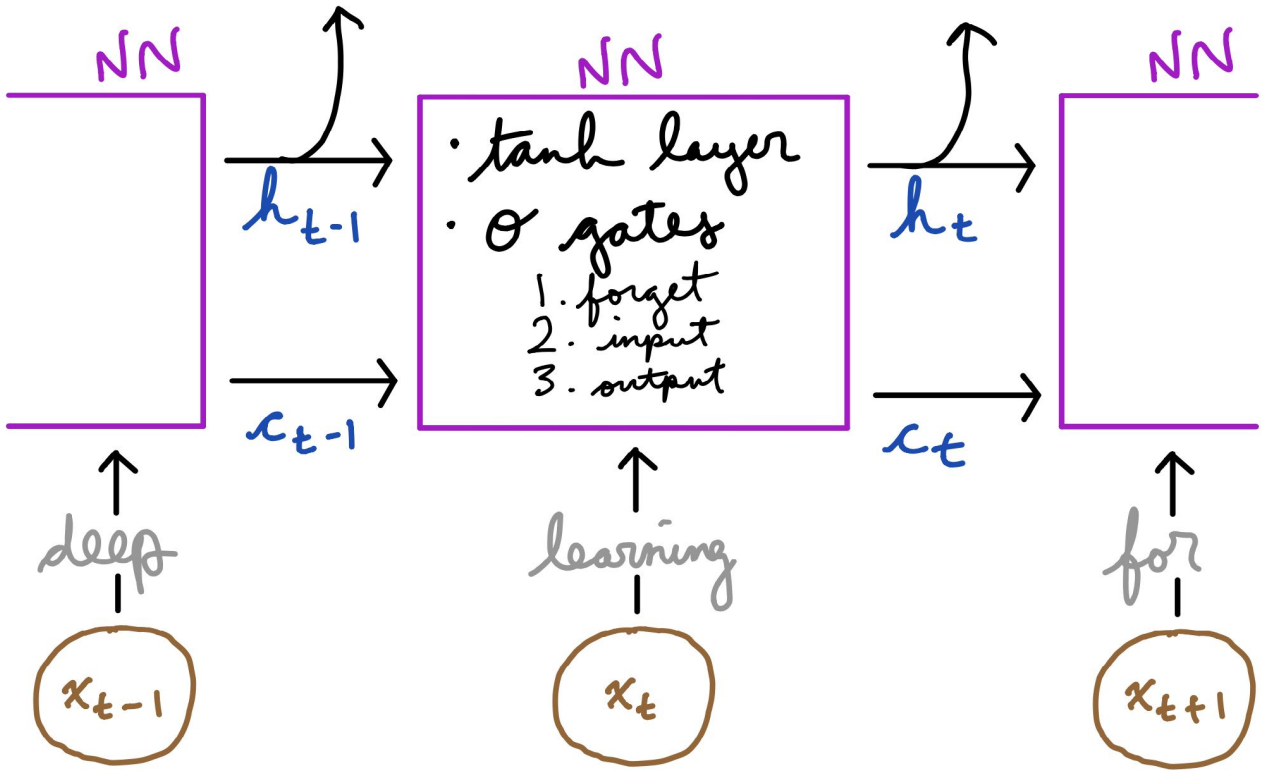
Recurrent Neural Networks



Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
   - Recurrent Neural Networks
   - LSTMs *(reference NLP LiveLessons 4.3 and 4.4)*
   - Bi-Directional LSTMs
   - Stacked LSTMs
   - Parallel Network Architectures

LSTM



Pearson

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
   - Recurrent Neural Networks
   - LSTMs
   - Bi-Directional LSTMs *(reference NLP LiveLessons section 5.1)*
   - Stacked LSTMs
   - Parallel Network Architectures

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
   - Recurrent Neural Networks
   - LSTMs
   - Bi-Directional LSTMs
   - Stacked LSTMs *(reference NLP LiveLessons 5.2)*
   - Parallel Network Architectures

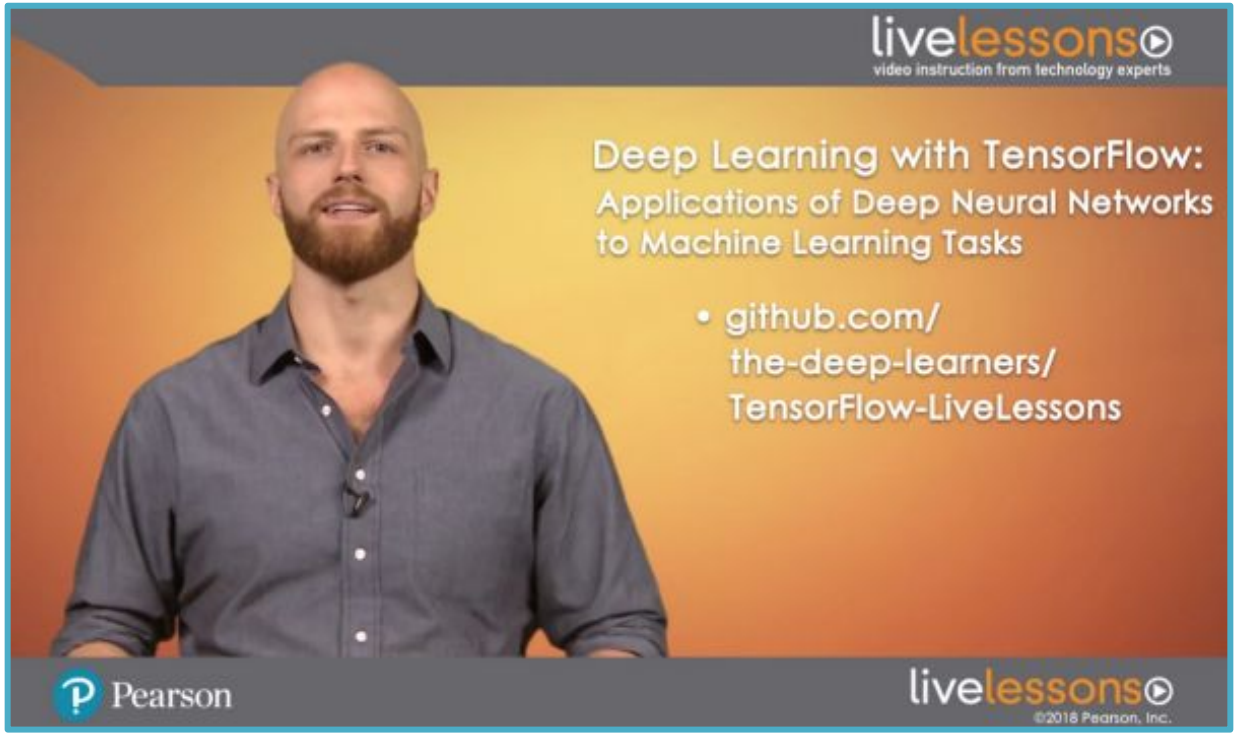
Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
   - Recurrent Neural Networks
   - LSTMs
   - Bi-Directional LSTMs
   - Stacked LSTMs
   - Parallel Network Architectures *(reference NLP LiveLessons 5.3)*

POLL

What other Deep Learning topic interests you most?

- CNNs and Machine Vision
- Generative Adversarial Networks
- Reinforcement Learning
- TensorBoard
- Something Else



Search *"Deep Learning with TensorFlow"* or *"Jon Krohn"* in Safari

Search *"Deep Reinforcement Learning and GANs"* or *"Jon Krohn"* in Safari

Staying in Touch

twitter.com/JonKrohnLearns

medium.com/@jonkrohn

linkedin.com/in/jonkrohn

(with message mentioning today's Live Training)

PLACEHOLDER FOR:

5-Minute Timer

PLACEHOLDER FOR:

10-Minute Timer