



Data Science Intro for HackExtend

Dalya Gartzman

goo.gl/5PEv8A

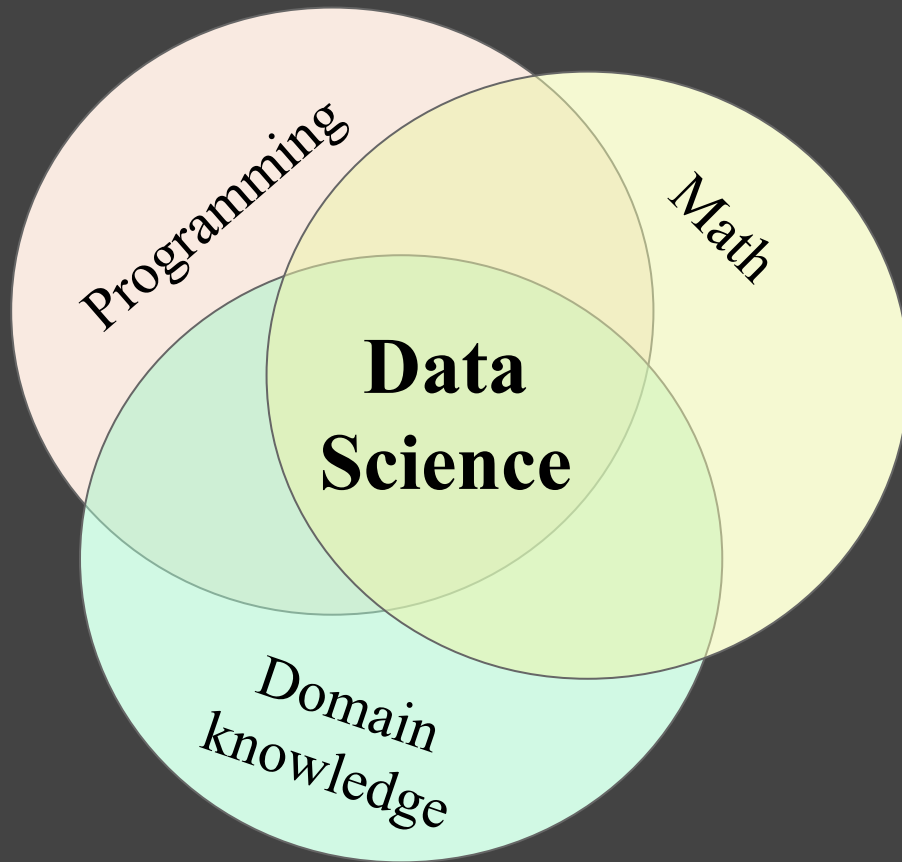
PART I - The Intro to the Intro

PART I - The Intro to the Intro

1. what is data science?
2. what is machine learning?
3. what is a neural network?

What is Data Science?

What is Data Science?



What is Data Science?



~~—What is Data Science?—~~

What do I need in my backpack
for a cool Data Science project?

~~What is Data Science?~~

What do I need in my backpack
for a cool Data Science project?

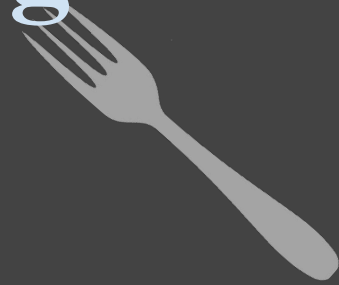
Programming

Basic Data
Science
Concepts

A Cool Idea



What is machine learning?



What is machine learning?

human learning:

shape(●) = ?



What is machine learning?

human learning:

If (# \sqcap) = 0

return circle

If (# \sqcap) = 4

return square

shape(\bullet) = ?



What is machine learning?

human learning:

If (# \sqcap) = 0
return circle
If (# \sqcap) = 4
return square


shape(\bullet) = ?


machine learning:

shape(\bullet) = ?

What is machine learning?

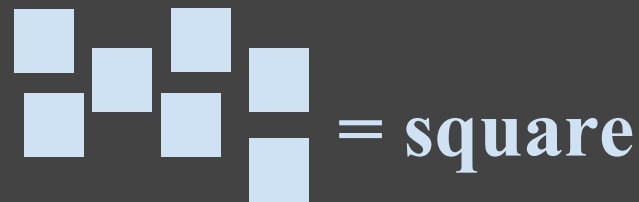
human learning:

If (# ) = 0
return circle

If (# ) = 4
return square

shape() = ?

machine learning:



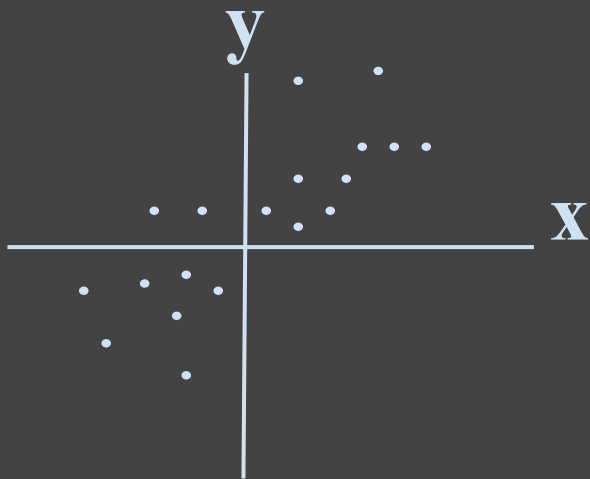
shape() = ?

What is a neaural network?



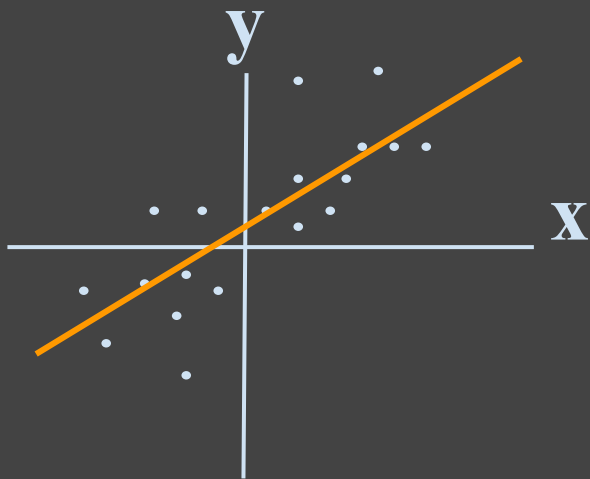
What is a neural network?

neuron:



What is a neural network?

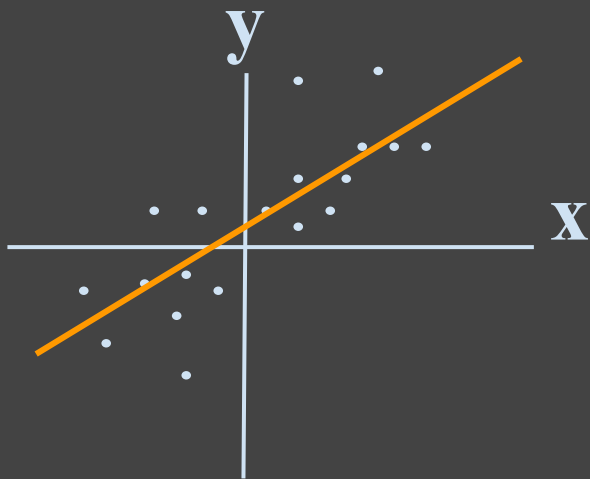
neuron:



$$y=f(x)$$

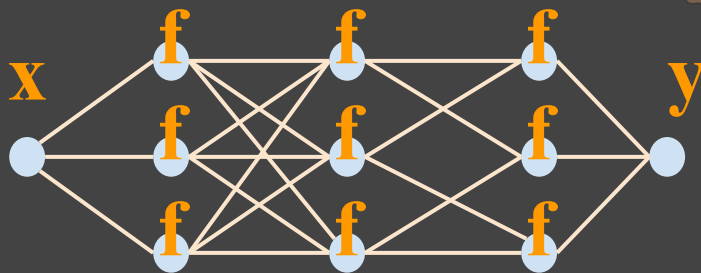
What is a neural network?

neuron:



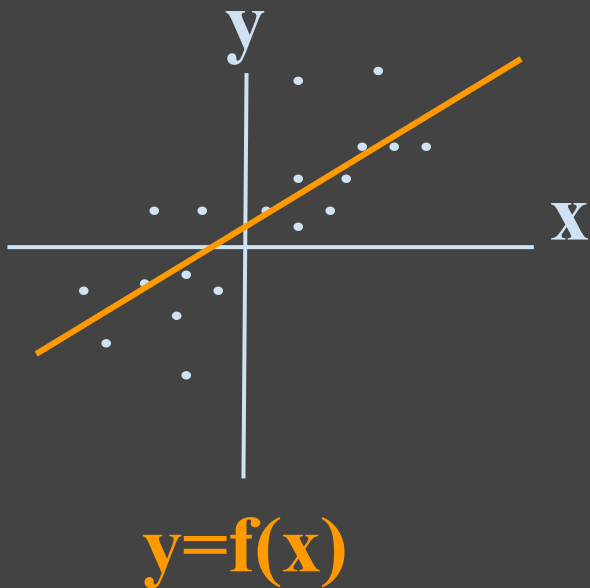
$$y=f(x)$$

neural network:

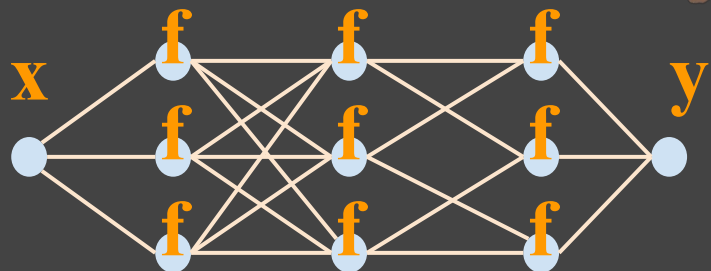


What is a neural network?

neuron:



neural network:



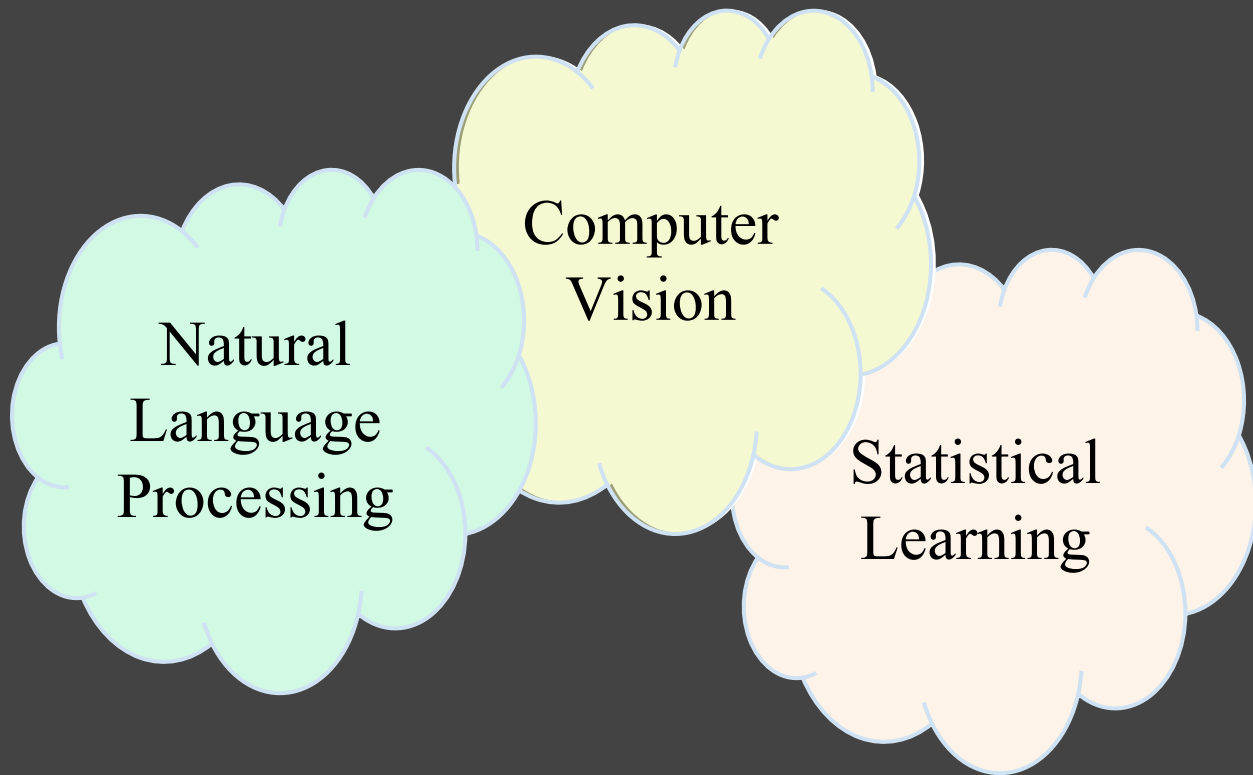
- architecture
- what are “f”?
- more stuff...

PART I Recap



PART II - so what can I do?

PART II - so what can I do?

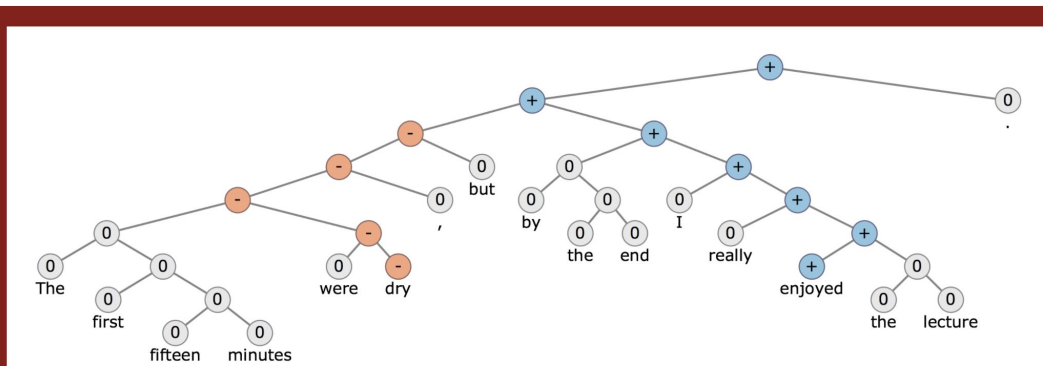


Natural Language Processing

Natural Language Processing

1. how i started

CS224n: Natural Language Processing with Deep Learning



[Syllabus \(Slides, etc.\)](#)

[Office Hours](#)

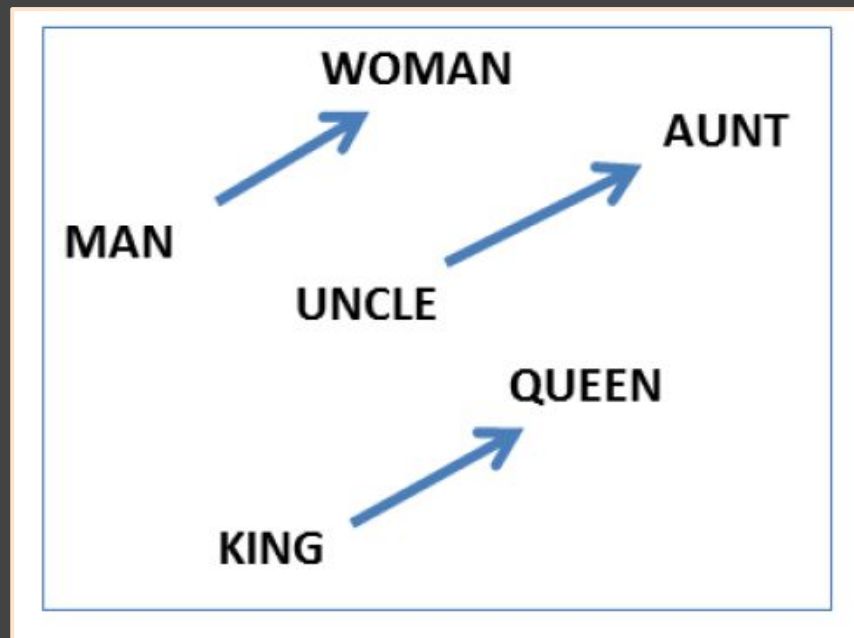
[Stanford Lecture Videos](#)

[Piazza forum](#)

[Project page](#)

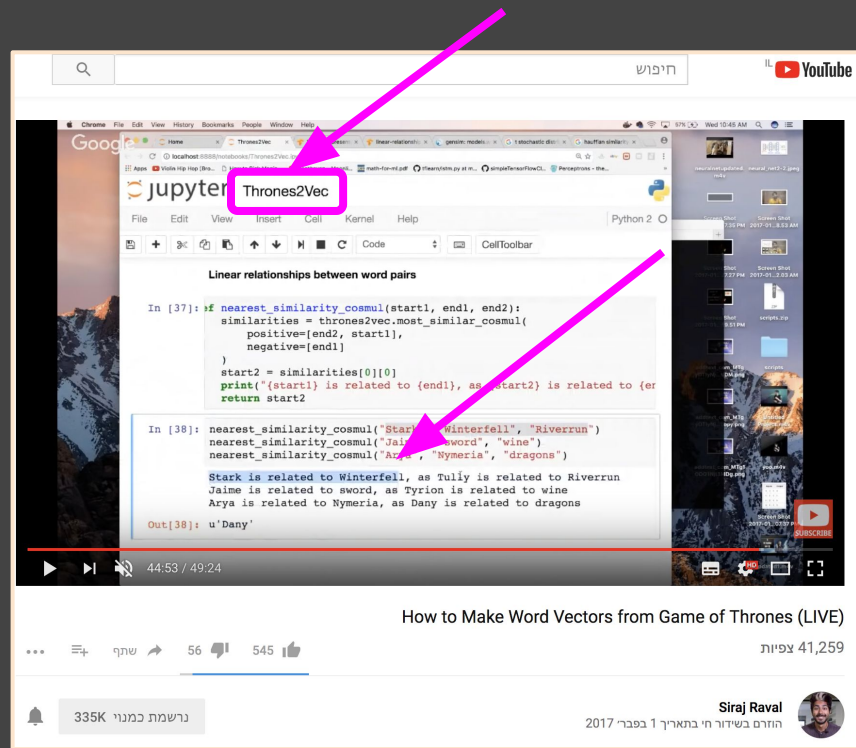
Natural Language Processing

1. how i started
2. what i learned



Natural Language Processing

1. how i started
2. what i learned
3. inspiration



The screenshot shows a Jupyter Notebook titled "Thrones2Vec" running Python code to calculate word similarities using cosine similarity. The code defines a function `nearest_similarity_cosmul` and applies it to various word pairs from Game of Thrones. The output shows the similarity between "Stark" and "Winterfell" is 0.0, and the similarity between "Jaime" and "sword" is 0.0. The output also shows the similarity between "Arya" and "Nymeria" is 0.0, and the similarity between "Stark" and "Winterfell" is 0.0.

```
Linear relationships between word pairs

In [37]: nearest_similarity_cosmul(start1, end1, end2):
similarity = thrones2vec.most_similar_cosmul(
    positive=[end2, start1],
    negative=[end1]
)
start2 = similarity[0][0]
print("{}(start1) is related to {}(end1), as {}(start2) is related to {}(end2)".format(start1, end1, start2, end2))
return start2

In [38]: nearest_similarity_cosmul("Stark", "Winterfell", "Riverrun")
nearest_similarity_cosmul("Jaime", "sword", "wine")
nearest_similarity_cosmul("Arya", "Nymeria", "dragons")

Stark is related to Winterfell, as Tully is related to Riverrun
Jaime is related to sword, as Tyrion is related to wine
Arya is related to Nymeria, as Dany is related to dragons

Out[38]: u'Dany'
```

How to Make Word Vectors from Game of Thrones (LIVE)

41,259 צפיות

נרשמת כמנוי 335K

הודום בשידור חי בתאריך 1 בפבר 2017

Siraj Raval

Natural Language Processing

1. how i started
2. what i learned
3. inspiration
4. implementation

Ender2Vec

```
In [43]: nearest_similarity_cosmul("Ender", "Valentine", "Bean")  
# interesting: http://enderverse.wikia.com/wiki/Suriyawong  
Ender is related to Valentine, as Suriyawong is related to Bean
```

```
Out[43]: 'Suriyawong'
```

```
In [44]: nearest_similarity_cosmul("Peter", "Valentine", "Locke")  
# amazing!!  
Peter is related to Valentine, as Demosthenes is related to Locke
```

```
Out[44]: 'Demosthenes'
```

```
In [61]: nearest_similarity_cosmul("run", "slow", "fight")  
run is related to slow, as survive is related to fight
```

```
Out[61]: 'survive'
```

```
In [64]: nearest_similarity_cosmul("love", "hate", "formics")  
love is related to hate, as officers is related to formics
```

```
Out[64]: 'officers'
```

```
In [ ]:
```

Natural Language Processing

1. how i started => web.stanford.edu/class/cs224n/
2. what i learned => www.tensorflow.org/tutorials/word2vec
3. inspiration => www.youtube.com/watch?v=pY9EwZ02sXU
4. implementation => github.com/DalyaG/Ender2Vec

Computer Vision


Computer Vision

1. some background

CS231n: Convolutional Neural Networks for Visual Recognition

Spring 2018

Previous Years: [\[Winter 2015\]](#) [\[Winter 2016\]](#) [\[Spring 2017\]](#)

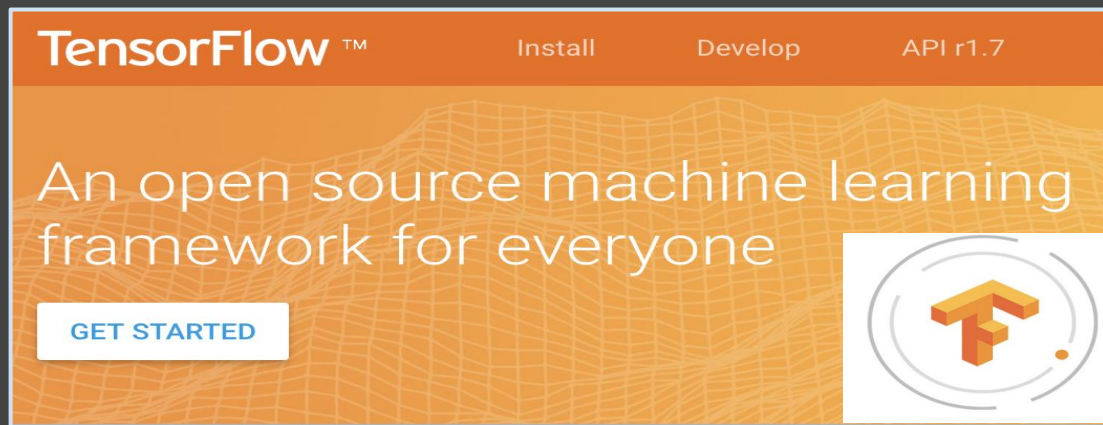


The diagram illustrates a convolutional neural network (CNN) architecture. It starts with an input image of a bird on the left. This image is processed through a series of layers, each represented by a grid of small images. These grids show the output of convolutional operations, where filters are applied to different parts of the input image to extract features. The layers progress from the input image to increasingly abstract feature maps. On the right side of the diagram, there is a vertical stack of colored bars representing the output probabilities for different classes. The classes listed are bird, deer, airplane, horse, and dog. The 'bird' bar is highlighted in green, indicating the network's prediction for the input image.

*This network is running live in your browser

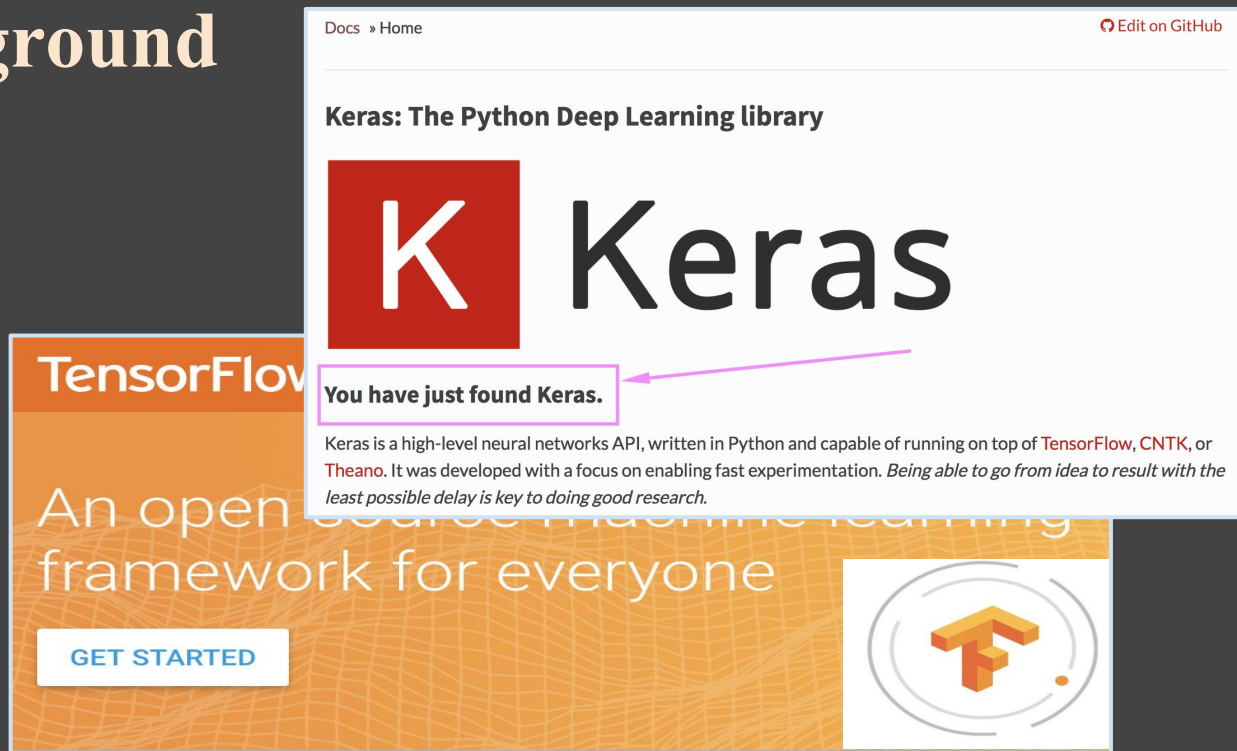
Computer Vision

1. some background
2. useful tools



Computer Vision

1. some background
2. useful tools



The image shows a screenshot of the Keras website. On the left, there is a large orange banner for TensorFlow with the text "TensorFlow" and "An open source machine learning framework for everyone". Below this banner is a white button that says "GET STARTED". On the right, there is a white box representing the Keras website. At the top of this box, it says "Docs » Home" and "Edit on GitHub". Below this, it says "Keras: The Python Deep Learning library". There is a red square with a white "K" and the word "Keras" in large black letters. Below this, there is a message in a pink box that says "You have just found Keras." with a pink arrow pointing to it. Below the message, it says "Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research." At the bottom right of the white box, there is a logo consisting of a stylized orange "T" inside a circle.

Computer Vision

1. some background
2. useful tools
3. don't sweat it

Usage examples for image classification models

Classify ImageNet classes with ResNet50

```
from keras.applications.resnet50 import ResNet50
from keras.preprocessing import image
from keras.applications.resnet50 import preprocess_input, decode_predictions
import numpy as np

model = ResNet50(weights='imagenet')

img_path = 'elephant.jpg'
img = image.load_img(img_path, target_size=(224, 224))
x = image.img_to_array(img)
x = np.expand_dims(x, axis=0)
x = preprocess_input(x)

preds = model.predict(x)
# decode the results into a list of tuples (class, description, probability)
# (one such list for each sample in the batch)
print('Predicted:', decode_predictions(preds, top=3)[0])
# Predicted: [(u'n02504013', u'Indian_elephant', 0.82658225), (u'n01871265', u'tusker', 0.1122357), (u
```


Computer Vision

1. some background
2. useful tools
3. don't sweat it

Usage examples for image class

Classify ImageNet classes with ResNet

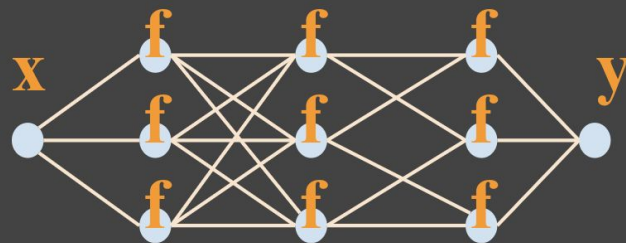
```
from keras.applications.resnet50 import ResNet50
from keras.preprocessing import image
from keras.applications.resnet50 import preprocess_input, preprocess_target
import numpy as np

model = ResNet50(weights='imagenet')

img_path = 'elephant.jpg'
img = image.load_img(img_path, target_size=(224, 224))
x = image.img_to_array(img)
x = np.expand_dims(x, axis=0)
x = preprocess_input(x)

preds = model.predict(x)
# decode the results into a list of tuples (class, score)
# (one such list for each sample in the batch)
print('Predicted:', decode_predictions(preds, top=3)[0])
# Predicted: [(u'n02504013', u'Indian_elephant', 0.82658225), (u'n01871265', u'tusker', 0.1122357), (u'
```

neural network:



- architecture
- what are “f”?
- more stuff...

Computer Vision

1. some background
2. useful tools
3. don't sweat it
4. here's one idea:

Computer Vision

1. some background
2. useful tools
3. don't sweat it
4. here's one idea:



Computer Vision

1. some background => cs231n.stanford.edu/
2. useful tools => www.tensorflow.org keras.io
3. don't sweat it => keras.io/applications

Statistical Learning

Statistical Learning


1. some background


Machine Learning

About this course: Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it. Many

[More](#)

Created by: Stanford University

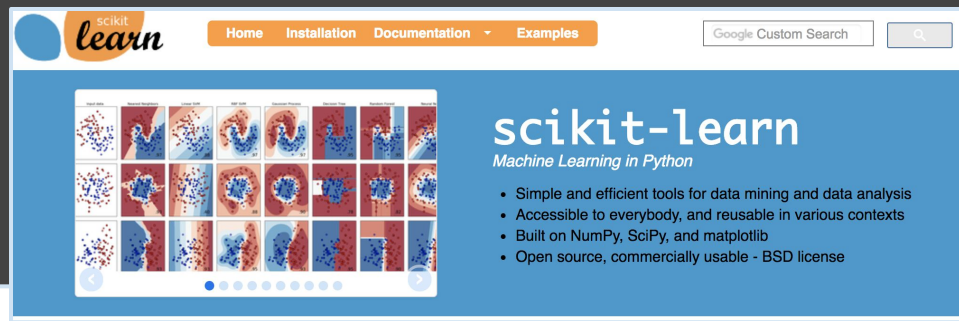




Taught by: [Andrew Ng](#), Co-founder, Coursera; Adjunct Professor, Stanford University; formerly head of Baidu AI Group/Google Brain

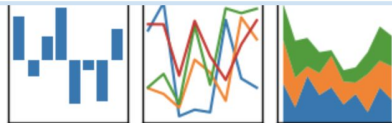
Statistical Learning

1. some background
2. useful tools



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



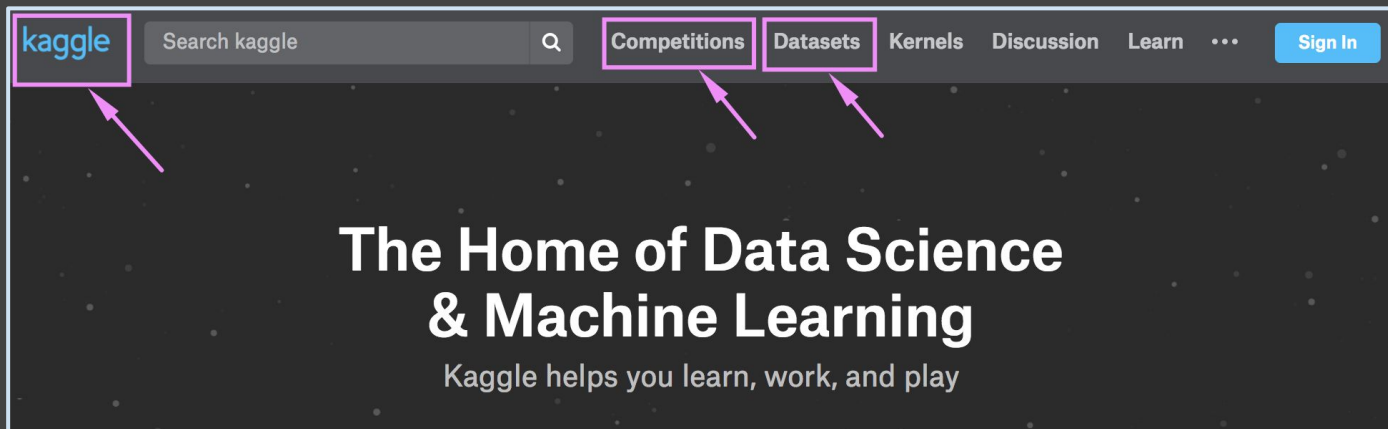
[home](#) // [about](#) // [get pandas](#) // [documentation](#) // [community](#) // [talks](#) // [donate](#)

Python Data Analysis Library

VERSIONS

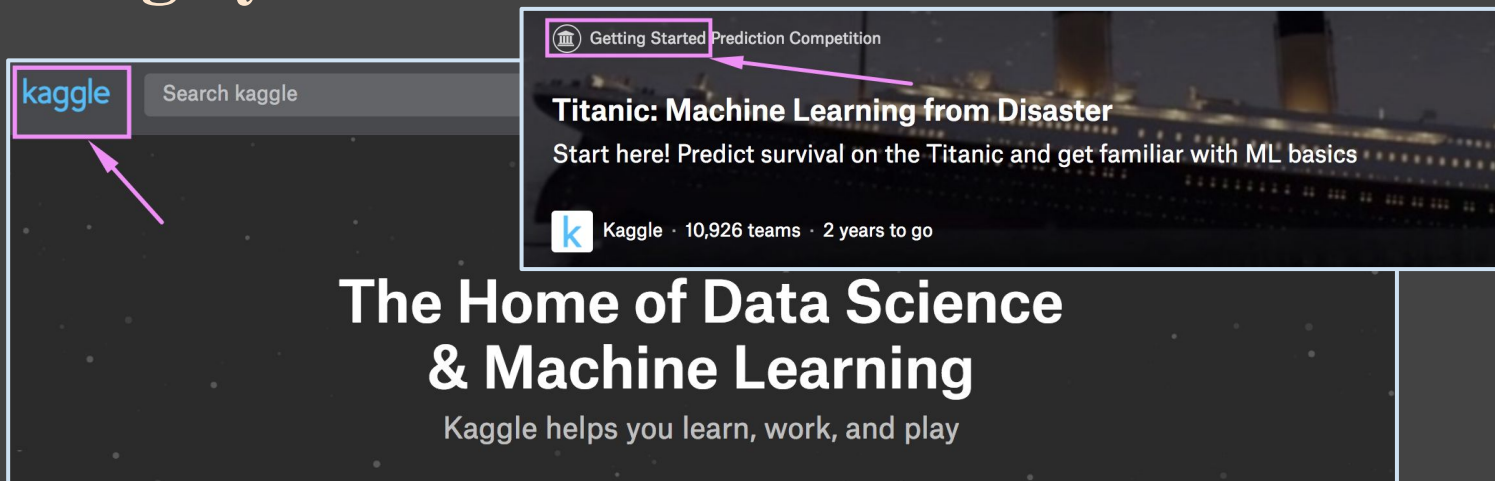
Statistical Learning

1. some background
2. useful tools
3. challenge yourself!



Statistical Learning

1. some background
2. useful tools
3. challenge yourself!



The image shows a screenshot of the Kaggle website. A pink box highlights the 'kaggle' logo, with a pink arrow pointing to it from below. Another pink box highlights the 'Getting Started' link in the top navigation bar, with a pink arrow pointing to it from the right. The main content area features a large banner for the 'Titanic: Machine Learning from Disaster' competition, which includes the text 'Start here! Predict survival on the Titanic and get familiar with ML basics' and 'Kaggle · 10,926 teams · 2 years to go'. Below the banner, the text 'The Home of Data Science & Machine Learning' is displayed, followed by the tagline 'Kaggle helps you learn, work, and play'.

kaggle Search kaggle

Getting Started Prediction Competition

Titanic: Machine Learning from Disaster
Start here! Predict survival on the Titanic and get familiar with ML basics

k Kaggle · 10,926 teams · 2 years to go

**The Home of Data Science
& Machine Learning**
Kaggle helps you learn, work, and play

Statistical Learning

1. some background
2. useful tools
3. challenge yourself!
4. here's some ideas:

★ Featured Dataset

Brewer's Friend Beer Recipes

Data on over 75,000 homemade beers



Statistical Learning

1. some background
2. useful tools
3. challenge yourself!
4. here's some ideas:

★ Featured Dataset

Brewer's Friend Beer Recipes

Data on over 75,000 homemade beers



How Much Did It Rain? II

Predict hourly rainfall using data from polarimetric radars

\$500 · 587 teams · 2 years ago

Statistical Learning

1. some background => www.coursera.org/learn/machine-learning
2. useful tools => pandas.pydata.org scikit-learn.org/
3. challenge yourself! => www.kaggle.com www.kaggle.com/c/titanic
4. here's some ideas => www.kaggle.com/jtrofe/beer-recipes
www.kaggle.com/c/how-much-did-it-rain-ii

PART II Recap



PART III - Team Up!

PART III - Team Up!

DATA

INSTALLATIONS

SOLUTIONS

TOOLS

ANALYSIS



Take Home Message



Thank you :)

Questions?

DalyaG@gmail.com