

Introduction to Data Science

...

Sign in at: tinyurl.com/bc2attendance

Main Topics:

- APIs
- K Means Clustering
- K Nearest Neighbors

Follow along with the Coding

- Go to github.com/cjleggett/hcs-bootcamp2
- Clone this repository to your computer
- The **demo.ipynb** file is where I'll be typing
- The finished product is the **API_Data_Science.ipynb** file

What is an API?

APIs (Application Programming Interface):

- Consist of a set of classes and functions
- Allow you to use functional code someone else has written.
- Always take the same input and produce the same output (allowing the API creator to edit code without affecting API users)
- Note: Some APIs require accounts/payment

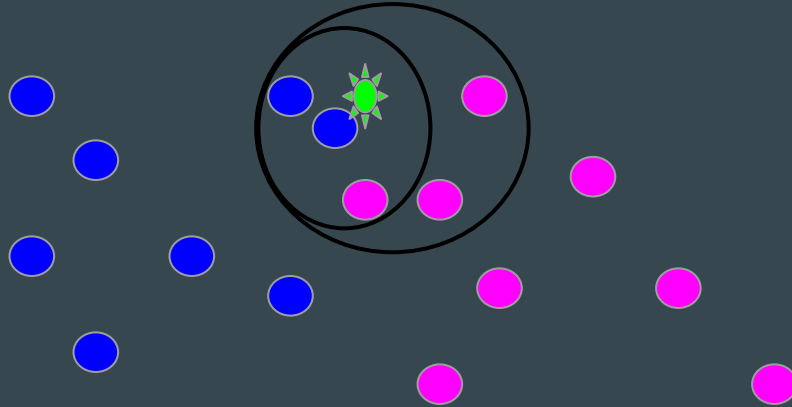
API Examples:

- Google Maps (paid)
- jservice.io (<http://jservice.io/>)
- lyrics.ovh
(<https://lyricsovh.docs.apiary.io/#reference/0/lyrics-of-a-song/search>)
- LOTS MORE:
(<https://github.com/toddmotto/public-apis#dictionaries>)

APIs in Python

Machine Learning

K Nearest Neighbors (Supervised)



Algorithm

1. Use dataset already split into clusters (This is why we call it 'supervised')
2. Find the k nearest data points to the new point (these are the nearest neighbors)
3. Classify the new point as whichever cluster contains the most nearest neighbors

Iris Dataset

- 3 types of Irises: Versicolor, Setosa, and Virginica
- Can we use KNN to classify Irises based on measurements?

Machine Learning in Python

Assignment (Groups of 1-3)

- Apply the machine learning technique we learned about today to a dataset of your choosing, and write an API where a client can input some info into the URL, and have a response returned
- Include a README file with a SHORT, INFORMAL explanation of what data you used and why you thought it was interesting.
- Ideas:
 - Find data on test scores and GPAs, and try to sort students into colleges
 - Find weather data and try to sort days into seasons based on data
 - Use data on sports teams to classify them as winning or losing
 - Anything else you think is interesting!

Assignment Tips

- Machine learning algorithms can work in many dimensions, but think about what you feel comfortable working with and visualizing before getting started.
- If you're interested in actually deploying your API to the web (so anyone can access it, not just your localhost) check out some free hosting services like [Heroku](#)
- Work in a group, and work on something interesting to you!

Helpful Links:

- Pandas documentation: <https://pandas.pydata.org/pandas-docs/stable/index.html>
- Overview of KNN Classification:
<https://www.kaggle.com/skalskip/iris-data-visualization-and-knn-classification>
- US Gov. Data: https://catalog.data.gov/dataset?res_format=CSV
- More csvs: <https://www.kaggle.com/datasets>
- Example code from bootcamp: <https://github.com/cjleggett/hcs-bootcamp2>

Submission Instructions:

- Submit at: tinyurl.com/bc2subs
- If you're working with a group, only 1 of you has to submit.
- Deadline: October 30, 11:59 pm