Lesson 22 – Hot Topics in ML

**Questions for Mentor:**

**NLP projects and startups to watch in 2019:**

* Natural Language Processing - Ability for a computer to understand human language
  + Component of AI

**Image Processing:**

* Very difficult, not as natural as our biological ability
  + Pose variability
  + Illumination variability
  + Intra-class variability
    - Same type of dogs, diff pics look different
  + Occlusion
    - Object being blocked by another object
* Convolutional neural networks
  + Convolutional layers stacked on top of each other
  + Signal in -> each is weighed -> decision made
  + Each layer takes a 3d volume, produces 3d volume with some smooth function that may or may not have parameters
  + Reduce parameters by reuse of spatial features across the image
  + Downsampling an image
    - Taking highest pixel value from chunk of image
* Fully convolutional networks
  + Goal: classify every pixel in an image
  + Difficulty: hard
  + Why?: when precise boundaries of objects matter (medical, driving)
    - Useful for fusing with other sensors (LIDAR)

**Recommendation Systems:**

* Content based – based on prior consumption, recommends similar items
* Collaborative filtering
  + Items entirely described by user ratings
  + Good for new discoveries
  + People who like scifi may also like fantasy
* Can use hybrid approach
* Similarity measures
  + Pearson correlation coef
  + Cosine similarity
* SVD – Singular Value Decomposition
* Map reduce
  + Programming model for very large datasets
  + Parallel distributed algorithms
  + Cluster framework
* Hadoop
  + Map reduce application
  + MrJob is a python interface in python
  + Will not run from ipython notebook
  + Launch the job
    - Python myscript.py < inputfile.txt > outputfile.txt
  + Map reduce can verify anagrams

**Network Analysis:**

* Examples
  + Social
  + Transportation
* Pathfinding
* Clustering
* NetworkX
  + API basics
  + Nodes and edges
* Undirected graphs
  + Facebook, users always follow each other
* Directed graphs
  + Twitter, users may not always follow each other, could be one way
* Network Visualization
  + Matriz plots
    - Nodes rows, columns are edges
  + Arc plots
    - Nodes are on axis of plot, edges are circular arcs over nodes
  + Circos plots
    - Like an arc plot, but are connected in a circle
  + Nxviz API
    - Can use to visualize above plots
  + Which nodes are important?
    - Number of neighbors