

# Amateur Adventures in Machine Learning

Nima Boscarino

# About me

- Instructor at Lighthouse Labs,  
Canada Learning Code
- Computing Science @ SFU
- Web Dev (full stack), with  
some other things sprinkled in
- @NimaBoscarino
- [nima.boscarino@gmail.com](mailto:nima.boscarino@gmail.com)



# What is this about?

- Machine Learning!
  - From a non-professional POV
- Broader picture:
  - Tips for learning new tech
  - “Work on a side project.” - They, 2012
    - What does that mean???

# What is Machine Learning?

- Loosely: software that gets better over time
- Leveraging data to make programs that adapt
- AI
  - replicating human functions (listening, reasoning, etc.)
  - Machine Learning is useful here, since human behaviours are generally pretty complicated
  - Can approximate behaviours with statistical models

# “What is” cont.

- Many methods for machine learning, today’s discussion involves neural networks
  - Slow to learn, fast to act!
- Basic idea for ML
  - Get a TON of labelled data
  - Partition into testing and training sets
  - Feed training set into model; model adapts to the training set
  - Evaluate performance using testing set
  - Hey presto! You got yourself a freshly-baked model

# Learning New Tech

- Generally difficult to find a place to start
- Picking a project makes some people anxious
- How do you avoid getting bored or frustrated?

# Phase 1

- Find a situation where you're "forced" to do something
  - e.g. take a course!
  - Machine Learning course at SFU
    - Required to complete assignments
      - Spam filter
      - Mushroom Recognizer



kaggle



# MyShroom

A neural network for recognizing mushrooms

# MyShroom

- Project for CMPT 419: Machine Learning
  - Taught by Greg Mori
  - Adapted an existing neural network (AlexNet)
  - Caffe
  - Had to construct the data set
    - 1600 mushroom pictures
    - ImageNet and Google Images
    - Tons of filtering and classifying by hand



# Phase 2

- Play time!
- Find silly ways of integrating the new tech into things you already do
- Focus on fun, not viability or seriousness
- Follow the memes
- Databased - the Hip Hop database



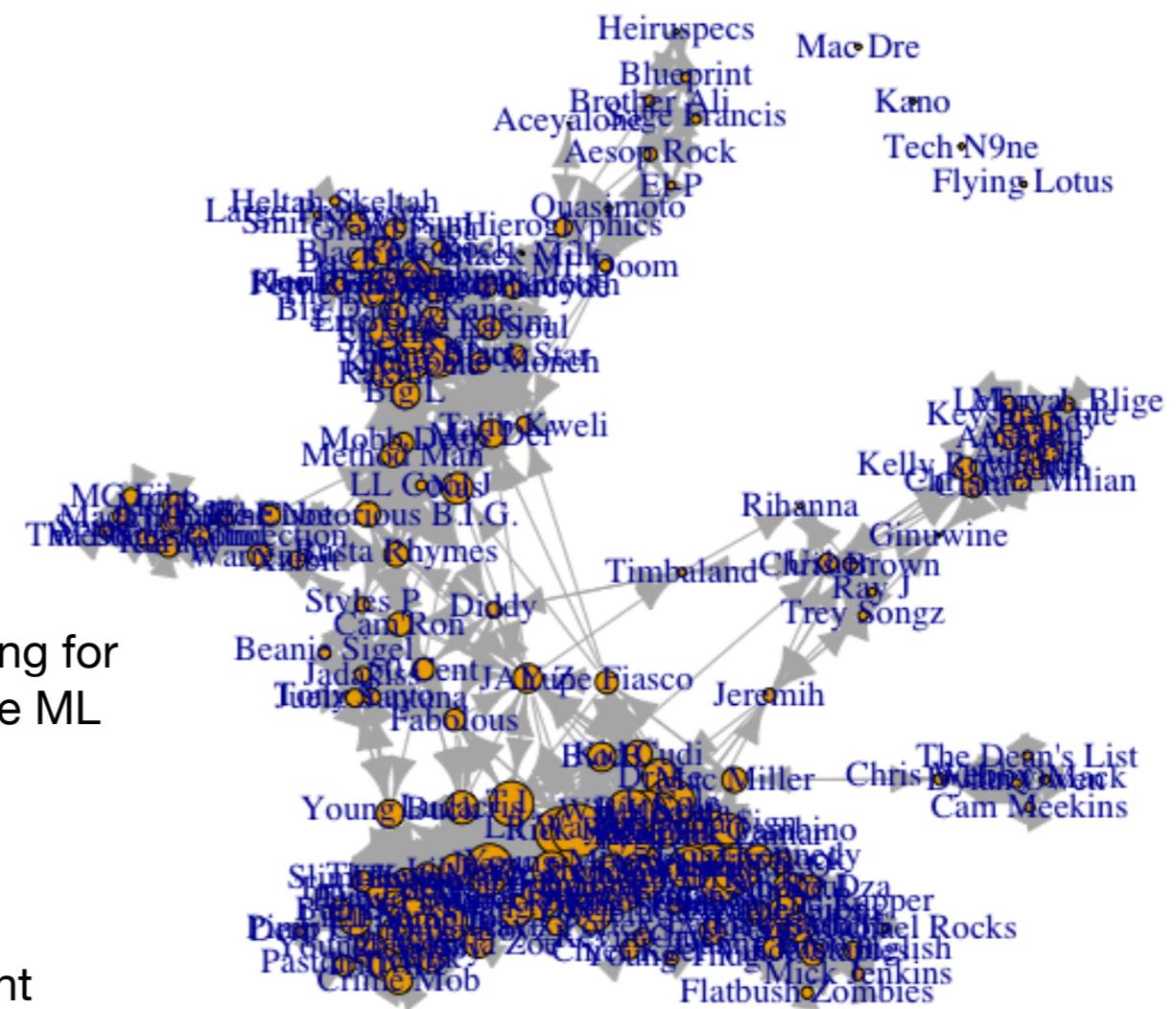


# DataBased

Scripts to build a Hip Hop database

# DataBased

- Main goals were...
    - MongoDB
    - Download scripts
    - Working with APIs
  - Inspiration from the Pudding article
  - Crossing Spotify, Genius, Last.FM
    - Different approach for each (e.g. scraping for Genius) “Hey... why don’t I try out some ML here?”
  - Andrej Karpathy
    - Unreasonable Effectiveness of Recurrent Neural Networks
    - Char-RNN



# A Jay Z Song

I don't lead your been the back the hards

When with the whore for me of the beat in the weal my rapper

I the house me my can the back that shit

I got she like a life the fuck my shit the shit

I cop the scrip my can you want me

Shit the hean the great back to be and the man to the back a man back

I got the for my life me out the can't come

So that they see shit your pop the sick and I strain

You was hold it you beat your brown the block you like the scare the still new then I'm niggas stude of the black  
the party

She wain the change the got the strop

How the been on the see to still the fleather

I know what the wear my life on the headers all the stuck the bropping the hood shit

I wat stop the world that the for that the house the can a back

You let you stop my dies what the change me

You don't a don't read to my game thing the for the boot to the shit

# Phase 3

- Find “serious” use cases
- Use your newfound skills to do something useful!
- Learn about real-life scenarios and pitfalls
- Hatfield wildlife camera
  - A situation where I saw a viable use for machine learning



07/15/2008 14:01



# Hatfield CC640

Text recognition for wildlife camera images

# Hatfield CC640

- Hatfield Consultants - environmental consulting
- Wildlife cameras with no EXIF data, needed to sort images
- Colleague tried out OCR (Tesseract) - no success!
- ML to the rescue...
- 1st try:
  - Find data set by scouring the internet
  - Write scripts to download, tile, classify
  - Train! (TensorFlow for Poets)
  - ... Didn't work! Overfitting!
- 2nd try:
  - Augment the data by warping (Keras)
  - Still didn't work on my images... Hmm...



# Recap

- “Just work on some side projects”
  - Find structured work
  - Integrate new tech into your existing activities (fun!)
  - Apply new knowledge in useful scenarios
- “Projects” arise organically as you play around
- Don’t stress about making “useful” projects
- Each of the phases takes time
  - There’s no need to rush!

# Resources

- Andrej Karpathy
  - Holy moly, so much stuff!
  - “The Unreasonable Effectiveness of Recurrent Neural Networks”
- Kaggle - lots of data sets!
- I have a couple youtube videos... haha
- DeepLoco - ML for 3D physics-based walking animation
- OpenAI Gym
- Marl/O by SethBling
- The Pudding ([pudding.cool](http://pudding.cool)) - Not ML, but interesting data stuff

# Thanks!

- @NimaBoscarino
- [nima.boscarino@gmail.com](mailto:nima.boscarino@gmail.com)

