

Data Science Methods for Clean Energy Research

Week 10 L2: This Is the End
March 8, 2017



Outline

- > Stuff
- > Things



Topics last time

> *In which I made your hearts race with a dazzling tale of unsupervised learning and intrigue*

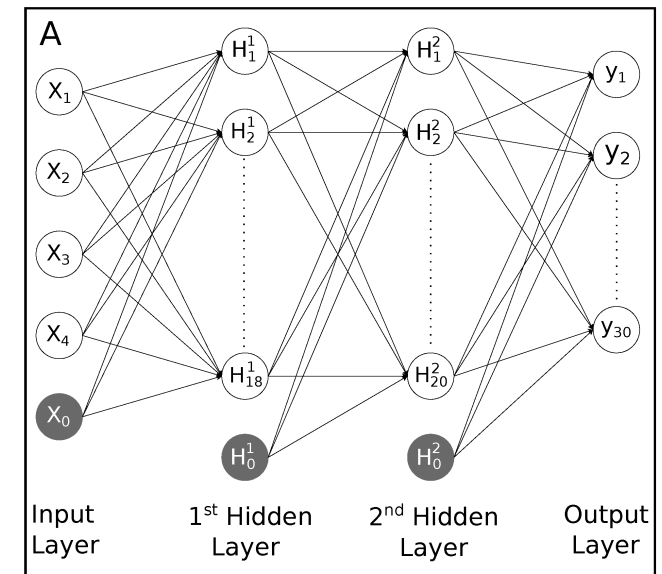
Big picture concepts:

- *The brains of DIRECT cohort 1 are fully cooked and ready to come out the oven*



What is a neural network?

- > Hidden layer neural networks are a supervised machine learning method
- > Make predictions of one (or more) responses (Y) given a set of features (X)
- > Inference about the relationships is essentially impossible, but predictive accuracy can be quite good
- > Require large training and validation sets. Training cost can be high!
- > Huge amount of tutorials, free information, literature, etc available online to learn more
- > See, in particular, Elements of Statistical Learning CH11



Neural networks in python

- > **Sklearn neural net:** no GPU implementation and not viewed as ready for “large scale applications”. Probably a great learning/teaching tool!
- > **Keras:** libraries built for Python, large user base and seen to be faster and well supported. Uses the TensorFlow library (a graph system for your ML data)
- > **Lasagna (nolearn):** libraries built for Python, large user base and well supported. Uses Theano (a different graph system for your ML data)



Where to learn more about time-series machine learning (when you are ready)

- > **Basic concepts in ML for time series data:**
 - You can apply a correction for either drift or periodic seasonality in your data
 - > This solves the problem of analyzing means and variances of your data
 - > This DOES NOT solve the problem that your data can be oversampled (e.g., sampling too much during correlated periods!)
- > **Decent sets of notes online:**
 - <https://www.analyticsvidhya.com/blog/2015/12/comprehensive-tutorial-time-series-modeling>
 - <https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>



Time series ML in python

- > I couldn't find an obvious, well-supported, "TIME SERIES FOR MACHINE LEARNING" Python package
- > Seems like it is more of a niche area
- > There are many examples online, and people often write their own code to condition the data (as in last slide) and then use established ML methods
- > Take care!



Capstone project process

- > You have just the list and title of the projects
- > Each project has a 1-2 page description (you saw the template for this already)
- > Today: feedback from you on how the projects should be selected
- > Friday (hopefully): release “project book” to you
- > Monday: share feedback w/DIRECT faculty and feedback from faculty
- > Monday: give you instructions for project selection



How should I assign you to projects?

- > More specifically: how would you deal with the situation of over-subscribed projects?
- > Can we come up with a list of priorities?
 - E.g.: “Every EE gets their 1st choice” , then “Everybody from PNNL wearing a blue shirt gets to work with a chemist” ← those are not good rules tho
- > Any other comments / suggestions ?



How I would start a project: If I were a **DIRECT** trainee

- > Meet with my team asap
- > Meeting 1 topics
 - Agree on timetable and work effort
 - > Agree on a minimum # of hours working together / week and set a schedule
 - > Agree on what you will do if people don't show up and are not contributing
 - > Discuss that if people don't show up it is not because they care but because they are busy and might not be as good as time management as you are
 - > Distribute this "charter" or "contract" via email and everyone reply-all that they agree



How I would start a project

> Meeting 1 topics

- Determine what you know and “don’t know” about the project based on the description
- Make a preliminary plan for how you will complete the project using your DIRECT class projects as an example
- Plan your first meeting with your project sponsor
 - > Rank order the things you need to know in terms of how essential they are to starting the project – you might run out of time!



How I would mentor a project: If I were a DIRECT sponsor

> Assume the following

- A high level of independence
- Teams will ask me if they need help
- Someone will tell me what to do unless it is more than a regular meeting and high-level supervision of progress
- Students will tell me if the project is interfering with their other obligations

> Aspire to the following

- Projects will lead to a successful deliverable including software and/or publication
- Students will reinforce existing skills/tools and get something important for their resume/jobs

