Data, Environment and Society: Lecture 26: Neural Networks

Instructor: Duncan Callaway
GSI: Salma Elmallah

December 3, 2019

Annoucements

- Thursday: Career panel!
- Poster presentation: 3-6pm, Dec 17, Kvamme Atrium in Sutardja Dai Hall 3rd floor.
  - Poster instructions in GitHub, here.
- Project due Dec 18 6a.

Last time

In words, each neuron...

- Takes a vector of values as inputs
- Creates a scalar from a linear combination of the vector entries
- Passes the resulting scalar through an "activation function"
- Outputs a single value from that activation function

Terminology analogies:

- Electrical signal from other cells $\rightleftarrows$ input
- Neuron $\rightleftarrows$ Activation function
- Electrical signal to other cells $\rightleftarrows$ output

Last time

In words, each neuron...

- Takes a vector of values as inputs
- Creates a scalar from a linear combination of the vector entries
- Passes the resulting scalar through an "activation function"
- Outputs a single value from that activation function

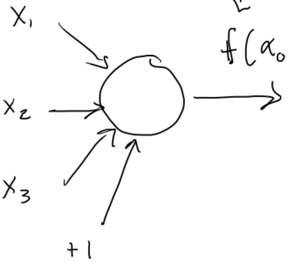
Terminology analogies:

- Electrical signal from other cells $\rightleftarrows$ input
- Neuron $\rightleftarrows$ Activation function
- Electrical signal to other cells $\rightleftarrows$ output

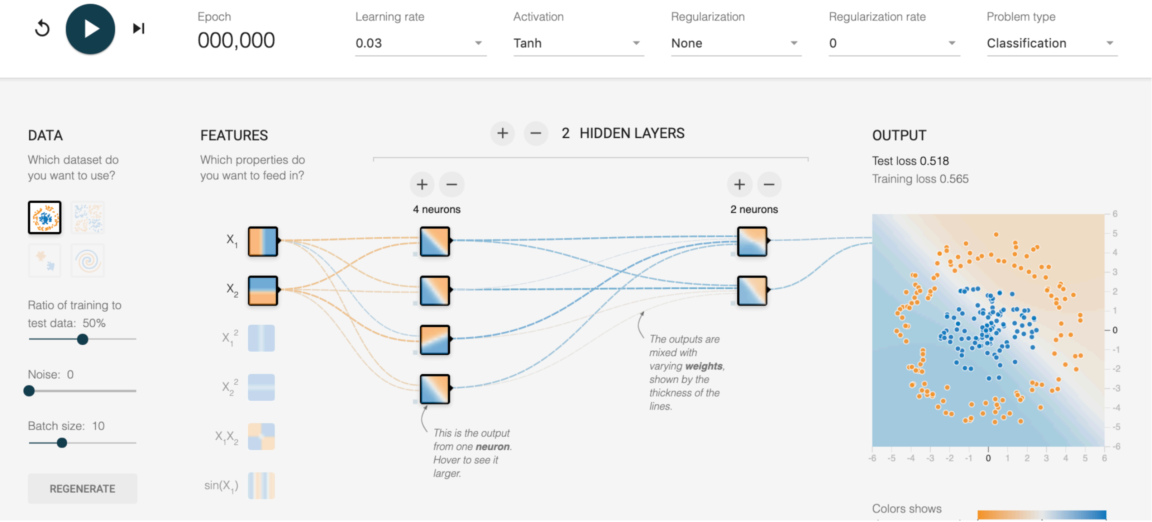
$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

activation f'n

$$f(\alpha_0 + \alpha^T x)$$



Glue neurons together, backpropogate, and off you go!



Today

Let's just talk about a few important and fun applications of neural networks.

Example resource allocation with NN

Combining satellite imagery and machine learning to predict poverty

Neal Jean[1,2,*], Marshall Burke[3,4,5,*,†], Michael Xie[1], W. Matthew Davis[4], David B. Lobell[3,4], Stefano Ermon[1]
+ See all authors and affiliations

Science 19 Aug 2016:
Vol. 353, Issue 6301, pp. 790-794
DOI: 10.1126/science.aaf7894

Article | Figures & Data | Info & Metrics | eLetters | PDF

Measuring consumption and wealth remotely

Nighttime lighting is a rough proxy for economic wealth, and nighttime maps of the world show that many developing countries are sparsely illuminated. Jean *et al.* combined nighttime maps with high-resolution daytime satellite images (see the Perspective by Blumenstock). With a bit of machine-learning wizardry, the combined images can be converted into accurate estimates of household consumption and assets, both of which are hard to measure in poorer countries. Furthermore, the night- and day-time data are publicly available and nonproprietary.

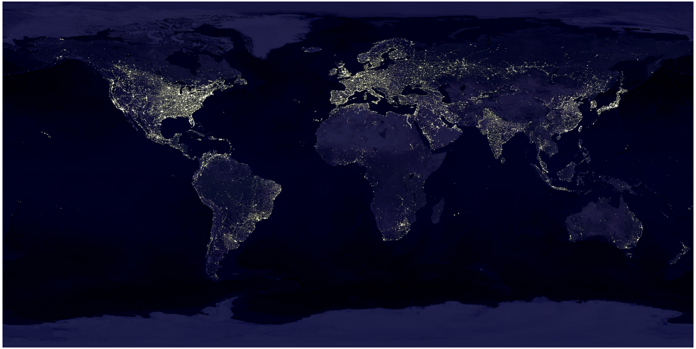
Science, this issue p. 790; see also p. 753

The challenge

- International aid agencies want information on where to spend effort
- Knowing where the poorest people are can inform these decisions
- But surveying a country's population is expensive, and most survey results are not in the public domain.
  - Some African countries do not run *any* surveys on wealth and poverty.
- Question: can one train a model that uses remote sensing data to *predict* poverty?

Night lights

- Some folks have used night satellite imagery to estimate spatial distributions of wealth
  - Basic idea: satellites can see lighting activity at night; People with more money use more night lighting.
- These methods are not accurate at extreme poverty income levels ($< \$1.90$ per person per day)
- There is little data on extreme poverty (see previous slide) so there isn't a large data set to train models with.



Jean *et al's* idea

1. Use *daytime* satellite imagery, not night time
   - Landcover type and structures, for example, ought to help to predict poverty.
2. Deal with the data paucity problem via "transfer learning"
   - The idea here is pretty simple: train a neural net on a well known set of images (ImageNet – labelled data from 1,000 categories, e.g. "boneshaker", "crutch", "miniature schnauzer")

   - its value isn't so much in finding schnauzers, but in being good at finding the low level features (edges, corners) common to many vision tasks

A bit more on Jean *et al's* transfer learning

3. ImageNet-trained model then “fine tuned” to use daytime satellite imagery inputs to produce features that predict nighttime light intensities outputs.
    - In doing this Jean *et al* are falling back on the idea of using night lights (or a prediction of night lights) to measure income and wealth
    - It's desirable to use night lights because it's a globally available data set and is a better proxy for economic activity than the daytime images would be.
4. Then build a ridge regression model:
    - Target: mean cluster-level values from *survey* data (where available)
        - Clusters are geographic areas approximately the size of a village.
    - Features: the corresponding image features extracted from the fine tuned ImageNet-trained model

Basic summary of process

- Old: night lights $\rightarrow$ predict wealth and income
- New: daytime imagery $\rightarrow$ predict night lights $\rightarrow$ predict wealth and income.

Wait – I thought night lights were a lousy proxy for economic activity at low income levels?

- Jean *et al* address (or try to address) this head on.
- Their claim is that because they're using a linear model (ridge regression) to map day time imagery to night lights, the model is going to be driven by light-economic relationships at higher income levels.
    - If they're lucky, the lower income relationship between *estimated* night lights and economic activity will be decent.

Results

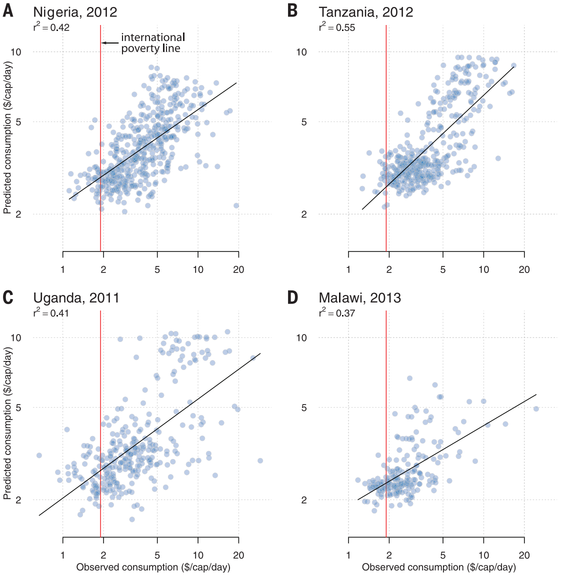
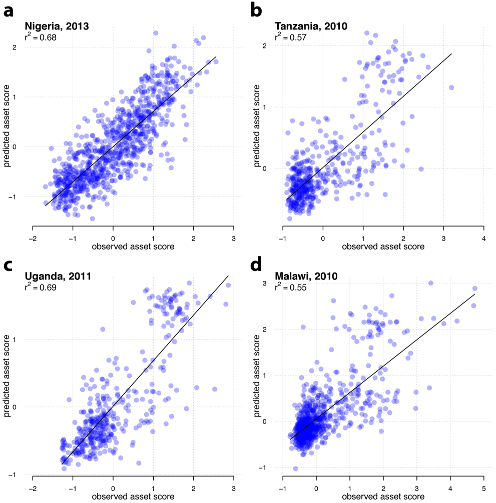


Figure S3: **Predicted cluster-level asset index from transfer learning approach (y-axis) c pared to DHS-measured asset index (x-axis) for 5 countries.** Predictions and reported r^2 va in each panel are from 5-fold cross validation. Both axes shown in log-scale. Black line is the best fit



External Validity

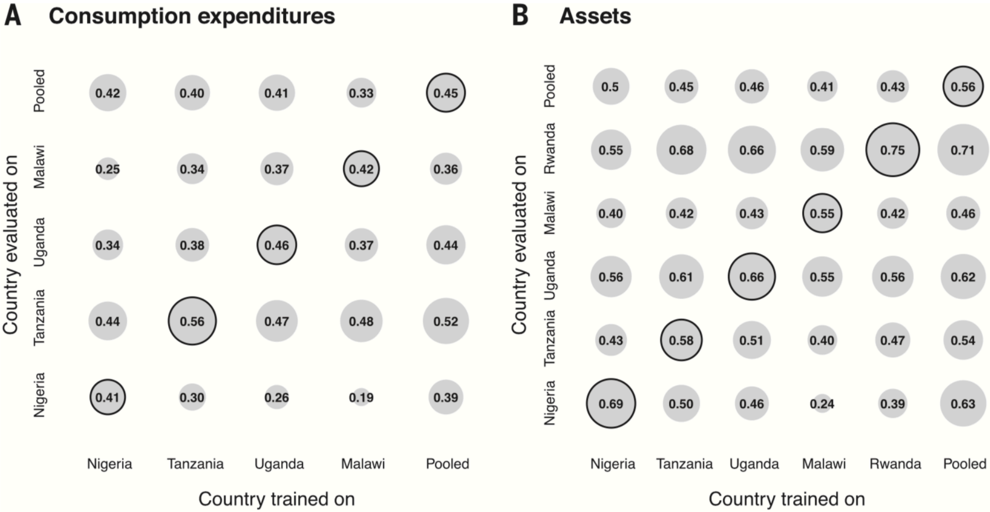


Fig. 5. Cross-border model generalization. (**A**) Cross-validated r^2 values for consumption predictions for models trained in one country and applied in other countries. Countries on *x* axis indicate where model was trained, countries on *y* axis where model was evaluated. Reported r^2 values are averaged over 100 folds

Another neural network story: Google Translate

Early in November, 2016, Google Translate produced the following:

> "Kilimanjaro is 19,710 feet of the mountain covered with snow, and it is said that the highest mountain in Africa. Top of the west, Ngaje Ngai in the Maasai language, has been referred to as the house of God. The top close to the west, there is a dry, frozen carcass of a leopard. Whether the leopard had what the demand at that altitude, there is no that nobody explained."

This is the translation of University of Tokyo Professor Jun Rekimoto's Japanese translation of the first paragraph of Hemingway's "The Snows of Kilimanjaro"

But something happened overnight:

One of these is Hemingway's original. One is the Google Translate version. Which is which?

"Kilimanjaro is a snow-covered mountain 19,710 feet high, and is said to be the highest mountain in Africa. Its western summit is called the Masai 'Ngaje Ngai,'' the House of God. Close to the western summit there is the dried and frozen carcass of a leopard. No one has explained what the leopard was seeking at that altitude."

"Kilimanjaro is a mountain of 19,710 feet covered with snow and is said to be the highest mountain in Africa. The summit of the west is called 'Ngaje Ngai' in Masai, the house of God. Near the top of the west there is a dry and frozen dead body of leopard. No one has ever explained what leopard wanted at that altitude."

But something happened overnight:

One of these is Hemingway's original. One is the Google Translate version. Which is which?

"Kilimanjaro is a snow-covered mountain 19,710 feet high, and is said to be the highest mountain in Africa. Its western summit is called the Masai 'Ngaje Ngai," the House of God. Close to the western summit there is the dried and frozen carcass of a leopard. No one has explained what the leopard was seeking at that altitude."

"Kilimanjaro is a mountain of 19,710 feet covered with snow and is said to be the highest mountain in Africa. The summit of the west is called 'Ngaje Ngai' in Masai, the house of God. Near the top of the west there is a dry and frozen dead body of leopard. No one has ever explained what leopard wanted at that altitude."

Are neural networks becoming more like humans? Or just better at imitating them?

- In this case, Google moved to a NN engine that translates whole sentences rather than phrases. In effect, it's just solving a much bigger classification problem than before.

A side note on using models trained on one data set to interpret another

A side note on using models trained on one data set to interpret another

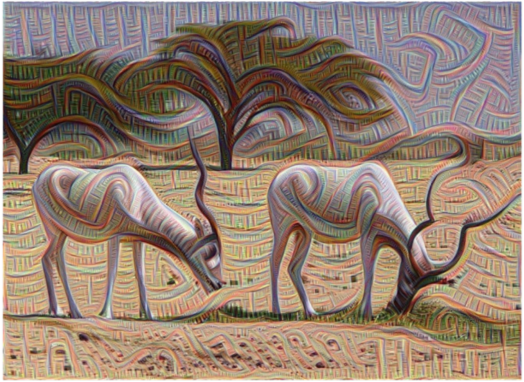
Image classification:

- Each layer of NN extracts higher and higher-level features of the image
- The final layer makes a decision on what the image shows.
- Example:
  - First layer might look for edges or corners.
  - Intermediate layers interpret basic features to identify components (wheel, cliff...)
  - Final layers assemble those into complete interpretations (bicycle, mountain range...)
- Google researchers played around with this:
  - Starting with an NN trained on ImageNet
  - Train the NN predict – or interpret – parts of new images
  - Look at intermediate layers of the interpretation

More information **here**; let's look at a few of their images now.

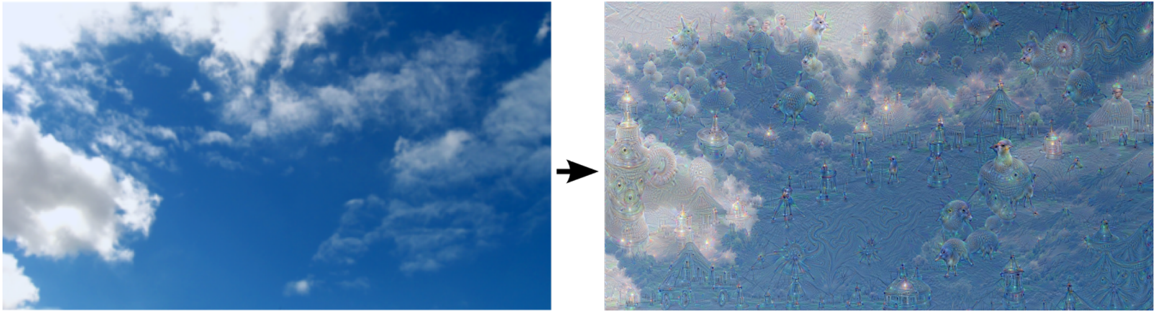
Example: Low layer interpretation

Here we see accentuation of strokes and colors.



Example: High layer interpretation

Here we see accentuation of specific kinds of features



A zoom in on the cloud images

"Admiral Dog!"

"The Pig-Snail"

"The Camel-Bird"

"The Dog-Fish"

A zoom in on the cloud images

"Admiral Dog!"

"The Pig-Snail"

"The Camel-Bird"

"The Dog-Fish"

Are these the dreams of neural networks?

- Reporters that covered the research liked to spin it that way
- I tend to think of this more pragmatically – these are just complex mathematical interpretations of images.
- BUT: I do think the images are beautiful and related to the way humans play with images
- This certainly pushes computing in the direction of art and spirituality, which I find scary and inspiring

Course wrapup: What do I hope you've learned?

Python

hw10.ipynb

Code Python 3

```
[ ]: NAME = "" # put your full name here
     COLLABORATORS = [] # list names of anyone you worked with on this homework.
```

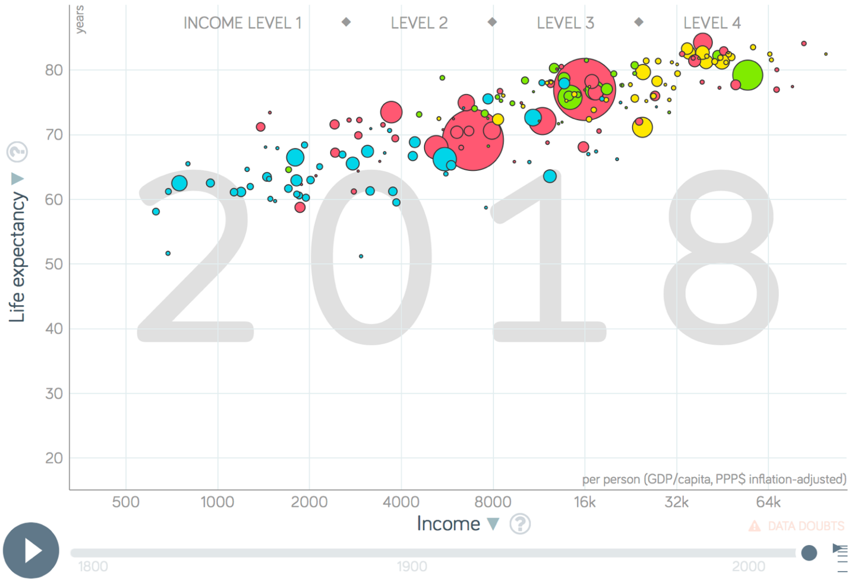
[ERG 131] Homework 10: Support Vector Machines

This homework will use support vector machines to classify CalEnviroScreen data. We will take gradual steps in this homework, starting from recalling key information from lectures and textbook, to creating our own classifiers. Throughout the homework, we'll learn about the intuition behind the Perceptrons and Maximal Margin Classifiers (MMC), then move on to learning about the intuition behind support vector machines (SVMs) and applying them to CalEnviroScreen data. The textbook reference here is ISLR 9.1-9.3.

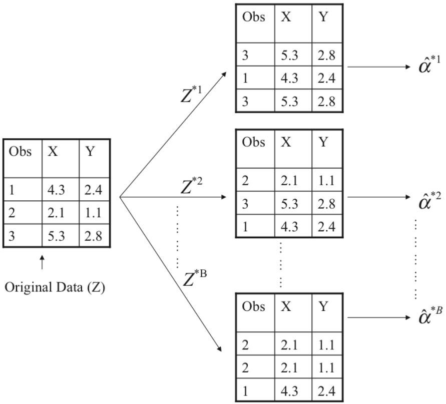
Table of Contents

Dependencies:

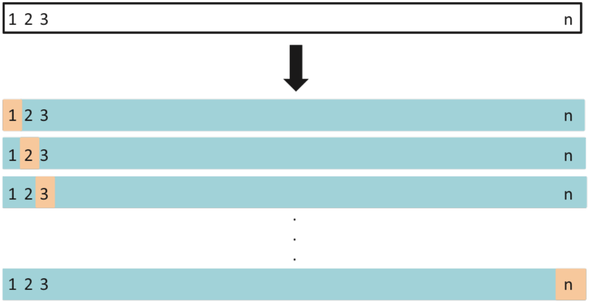
Visualization



Resampling



Obs X Y
1 4.3 2.4
2 2.1 1.1
3 5.3 2.8
Original Data (Z)
Z^{*1}
Z^{*2}
Z^{*B}
Obs X Y
3 5.3 2.8
1 4.3 2.4
3 5.3 2.8
Obs X Y
2 2.1 1.1
3 5.3 2.8
1 4.3 2.4
Obs X Y
2 2.1 1.1
2 2.1 1.1
1 4.3 2.4
$\hat{\alpha}^{*1}$
$\hat{\alpha}^{*2}$
$\hat{\alpha}^{*B}$



1 2 3 n
1 2 3 n
1 2 3 n
1 2 3 n
1 2 3 n

Model identification

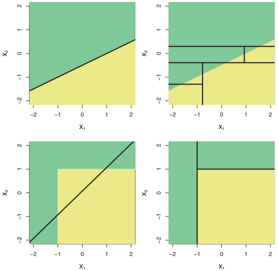
- Parameters fit by loss minimization
  - These determine the shape of the function you're fitting
- Hyperparameters fit by iteration and cross validation
  - These enable you to tune bias-variance tradeoff

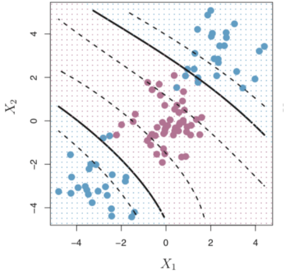
Zoom out: Train, (cross) validate and test

Using slightly different language, some definitions from from Brian Ripley, Pattern Recognition and Neural Networks, 1996, page 354:

- **Training set**: "A set of examples used for learning, that is to fit the parameters of the classifier."
  - For minimizing the loss function
- **Validation set**: "A set of examples used to tune the parameters of a classifier, for example to choose the number of hidden units in a neural network."
  - For choosing hyperparameters.
- **Test set**: "A set of examples used only to assess the performance of a fully-specified classifier."
  - For a final check – no more model fitting allowed here!

New tools for prediction: regression and classification



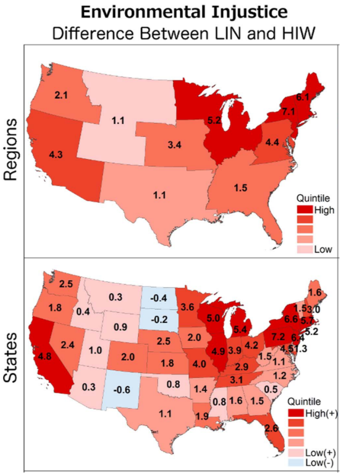


- Linear regression models: Ridge, Lasso, Elastic net

$$\min_{\beta} \sum_{i=1}^{N} (Y_i - X_i\beta)^2 + \lambda \cdot R(\beta)$$

- Decision trees
- Support vector machines

Inference, prediction and resource allocation in environmental contexts



- Prediction in space: what's happening in places where I have limited data?
  - Example: What are the NO2 concentrations where we don't have monitoring equipment?
  - Example: What are the land cover characteristics in a region?
- Prediction across communities: What's happening in communities where I have limited information?
  - Example: What communities in California have clean water?
- Prediction in time: What will happen next year, or tomorrow?
  - Example: will the air be clean tomorrow?

Next steps: What I hope for you

- You've now got basic skills for data manipulation and modeling and working in Python
- You've also got experience developing research questions and defining resource allocation problems
- You should feel proud of these skills when you present yourself to potential research advisers and employers
    - Environmental and Justice nonprofits and NGOs
    - Government regulators
    - Energy sector companies

  ...all these folks are looking for the skills you now have
- You're also ready to take more courses in the area: DS100, CS189, Stat 154.
- I also hope you'll keep in touch – tell me how you use these skills!

https://course-evaluations.berkeley.edu/