

Stat 412: Advanced Regression and Predictive Modeling

Course Description

Often we are interested in making inferences and predictions from data, either by (1) estimating particular meaningful parameters of models or (2) finding best fitting model that we can then manipulate to produce useful outputs such as predictions or counterfactual estimates. Focus on what is done when linear models are not appropriate and may produce misleading estimates. Generalized linear model and maximum likelihood methods as essential tools all statistics students should understand. Examination to shift gears to explore predictive modeling techniques that have been ubiquitous in machine learning literature [practice] in recent years, ~~with special attention to regularization and kernelized methods.~~

Course Pre-reqs

- Stat 402: Applied Regression
- Working knowledge of R (or similar)

Advanced Regression

simply put, regression is used to model the relationship between a dependent variable and one or more independent variables

often thought of as proposing a stochastic model that could have generated data, and then deriving methods to infer properties of the underlying generative mechanism.

Design-based Inference

- randomness comes from a controlled study design
- examples: randomized experiments, survey sampling
- reliable framework for causal inference
- not always possible

Model-based Inference

- how do we set parameters of a pre-specified model to best accord with the data?
- randomness comes from an assumed probability model
- good for predictive inference, characterizing data; can be used for causal inference only under difficult assumptions

*here we focus on model-based inference, but many projects involve aspects of both

Predictive Modeling

Constructing methods which predict well over some given data universe [specific dataset]

Prioritizing predictions, being less concerned with the underlying mechanisms generating the data, allowing for many different algorithms to be compared, and largely concerned with accuracies of predictions

Goals

- learn foundational modeling approaches (MLE)
- learn newer ubiquitous modeling techniques and the trade-offs of each
- prepare you to better learn the (many) things we can't cover
- improve your work, consulting, and research skills

Topics Part I - Maximum Likelihood, Generalized Linear Models

- binary outcomes: logit & probit as running examples
- likelihood theory, inference
- interpretation
- other common models: e.g. ordered, categorical, and count outcomes

Topics Part II - Machine Learning

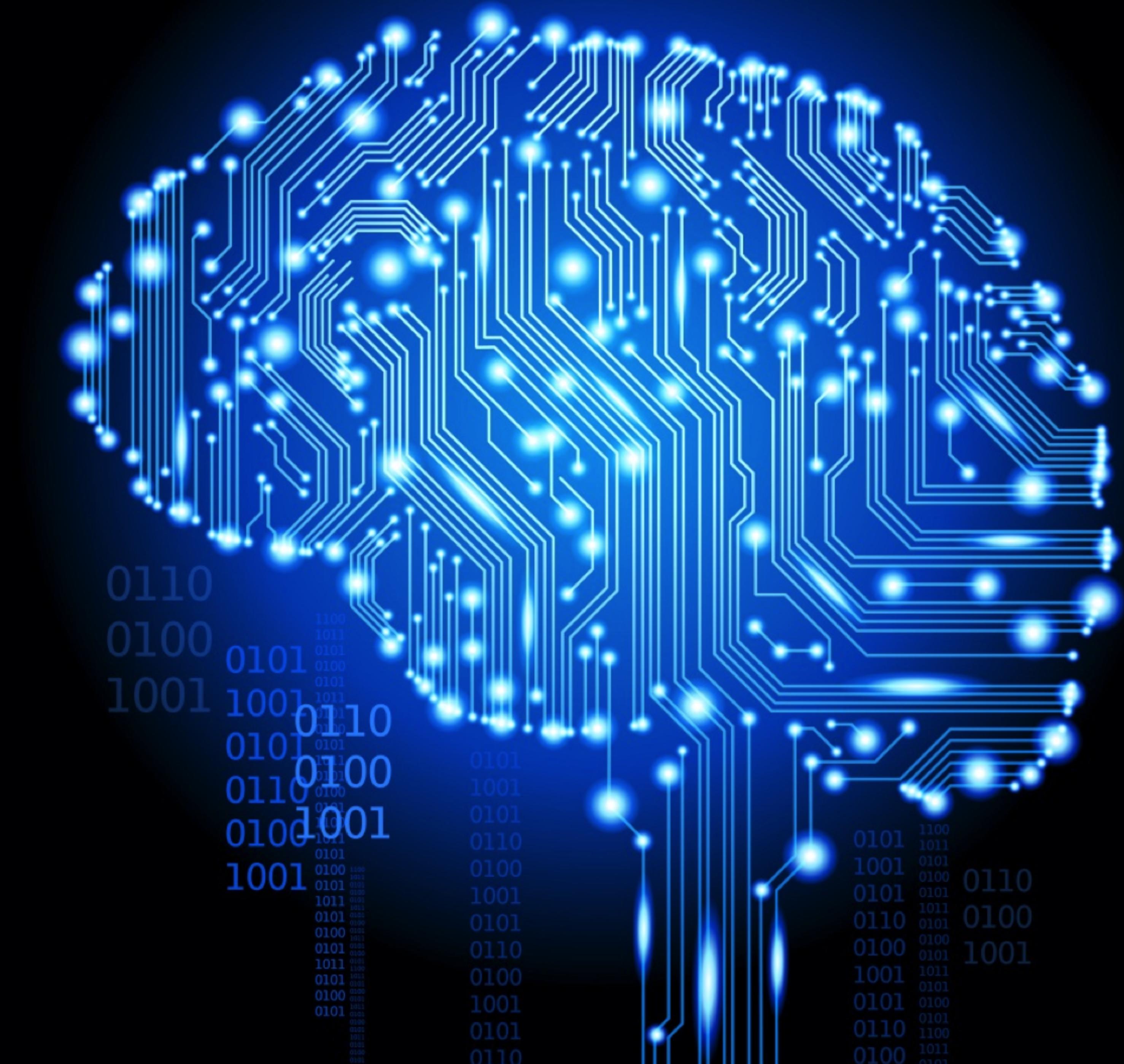
- focus on supervised rather than unsupervised learning
- a light dose of learning theory
- generalizability, cross-validation, out-of-sample testing
- regularization in linear models: ridge, lasso, elastic net,...
- tree-based methods: CART, Random Forest
- boosting and bagging
- kernels for operating in high-dimensional spaces, support vector machines

AN INTRODUCTION TO MACHINE LEARNING





```
mirror_ob.type = 'MESH' # and modifier_ob.type is 'CURVE'
mirror_ob.select = False # set to mirror_ob.type is 'CURVE' hope the other is a mesh
modifier_ob = bpy.context.selected_objects[0]
print("popped")
#mirror_ob
mirror_ob = bpy.context.active_object
print("Modifier object:" +str(modifier_ob.name))
#modifier_ob
modifier_ob = bpy.context.selected_objects[0]
print("mirror ob",mirror_ob)
print("modifier ob",modifier_ob)
t mirror modifier on modifier_ob
mirror_mod = modifier_ob.modifiers.new("mirror_mirror","MIRROR")
t mirror object to mirror_ob
mirror_mod.mirror_object = mirror_ob
operation == "MIRROR_X":
mirror_mod.use_x = True
mirror_mod.use_y = False
mirror_mod.use_z = False
operation == "MIRROR_Y":
mirror_mod.use_x = False
mirror_mod.use_y = True
mirror_mod.use_z = False
operation == "MIRROR_Z":
mirror_mod.use_x = False
mirror_mod.use_y = False
mirror_mod.use_z = True
```



Nate Langholz, Ph.D.
Statistician - Data Scientist



UCLA

Aaron Danielson, Ph.D.

Statistician - Data Scientist

SFU



Class Intros...



*feel free to use Jupyter notebooks and Python but lately I have been more inclined to exclusively use RStudio to do the topics covered in this course



<http://github.com>

<http://happygitwithr.com>

Course Info and Logistics

<https://github.com/natelangholz/stat412-advancedregression>



<https://www.kaggle.com>

we will be running a classroom
competition through kaggle inClass



effective communication skills

critical skill for statistician requiring
practice and feedback to develop

marketing

“For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt.All in all I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data”