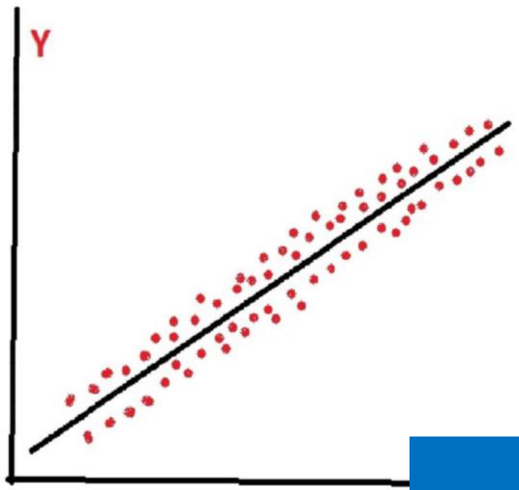
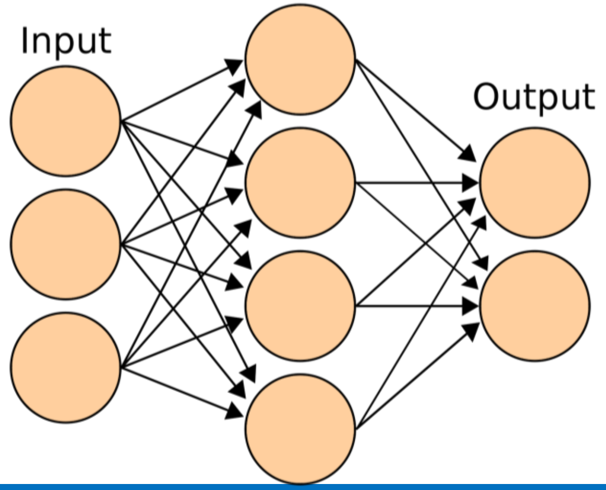
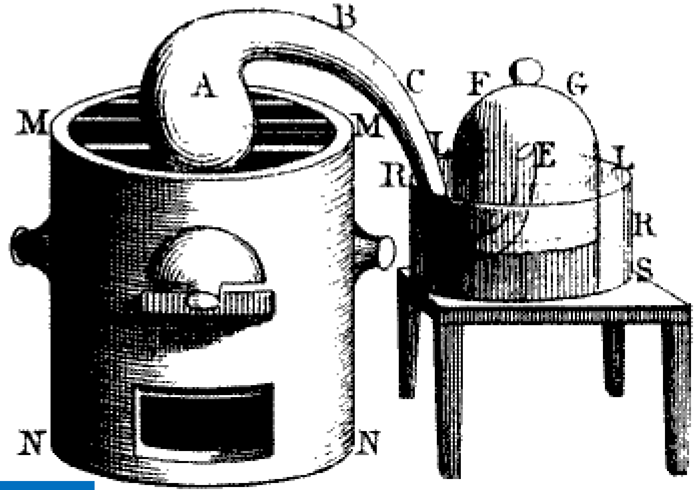




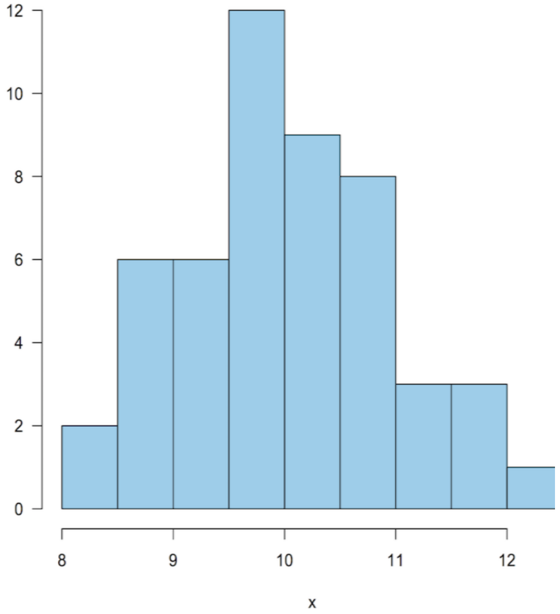


How We Use Data

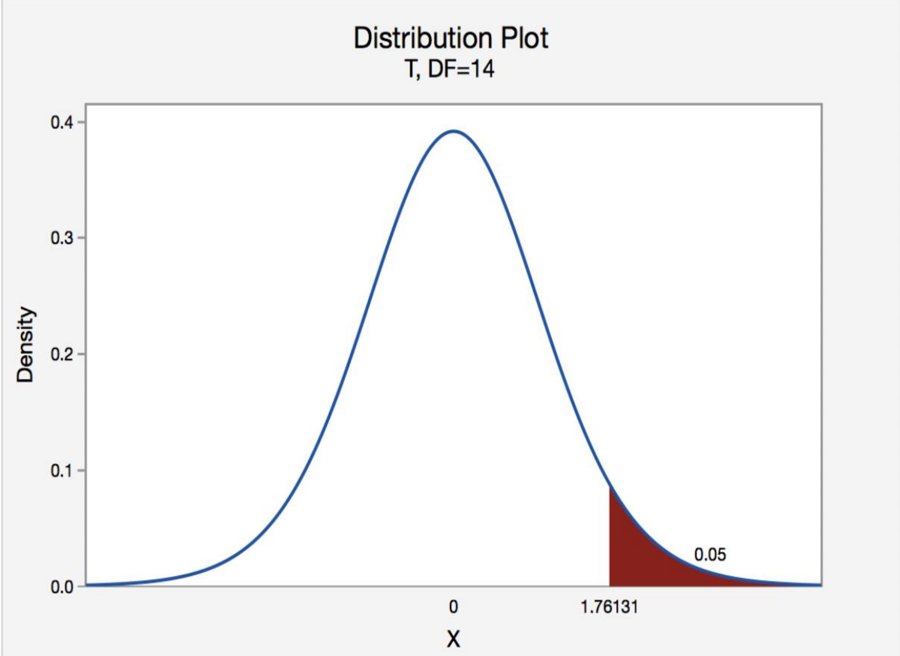
FSM Seminar

Spring 2024

Conor Rowan



$$\underline{\underline{X}} = \begin{bmatrix} u(x_1,t_1) & u(x_2,t_1) & \dots & u(x_N,t_1) \\ u(x_1,t_2) & u(x_2,t_2) & \dots & u(x_N,t_2) \\ \vdots & \vdots & \dots & \vdots \\ u(x_1,t_T) & u(x_2,t_T) & \dots & u(x_N,t_T) \end{bmatrix}$$



Background

- Data emerging as new paradigm in computational science
- Lots of hype and buzzwords ("machine learning," "data-driven," "AI," etc.)
- Less explored than traditional computational methods, so there are more opportunities to make contributions
- Hype tends to divide people into skeptics and optimists
- Data has revolutionized some fields, and not others

Motivation

- Explore the different ways in which data is used throughout the physical sciences
- Discuss some examples of machine learning and other data-driven methods
- Better understand what goes into successful applications of data-driven modeling
- **Persuade machine learning optimists to be more skeptical, and machine learning skeptics to be more optimistic**

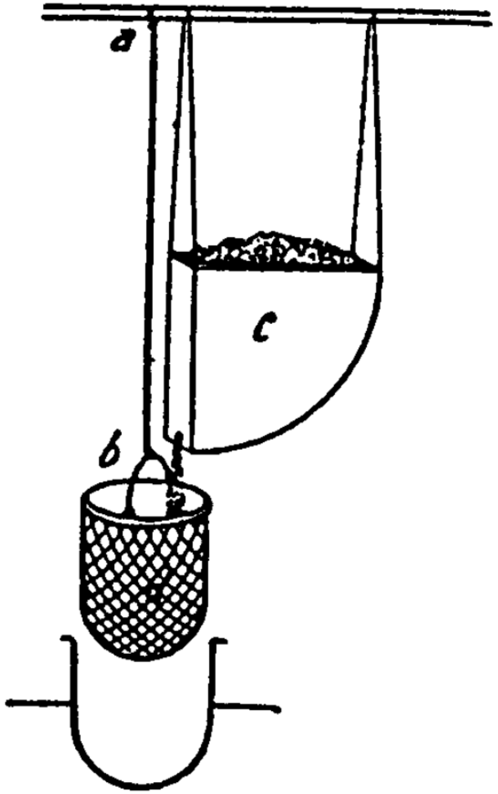
Model-free use of data

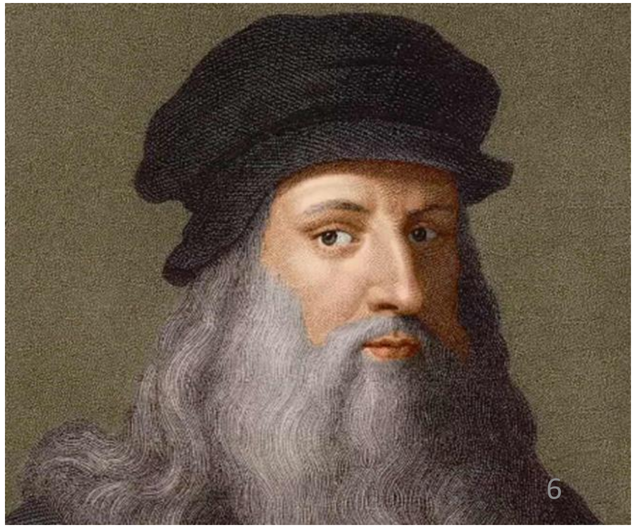
Qualitative conclusions

- Experimental work used to explore qualitative hypotheses
- No mathematical "post-processing" done on data obtained from experiments
- Conducted when underlying physical principles are not known, or partially known
- Characteristic of exploratory early scientific work, not common these days

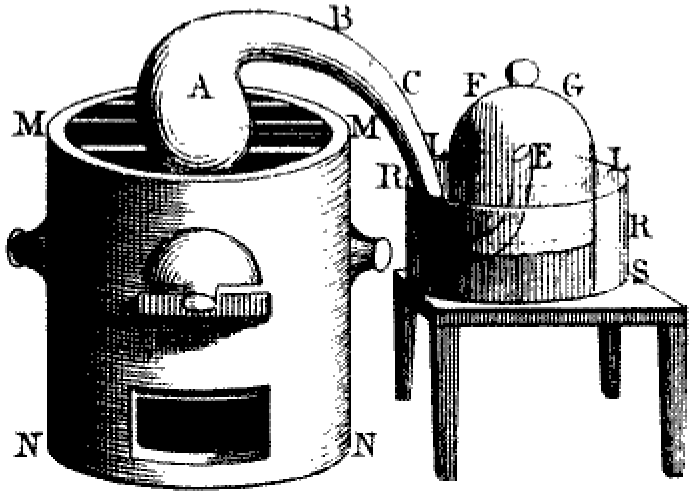
Da Vinci wire experiments

- Early experiments in engineering mechanics and fracture
- Pulls thin wires until they break
- Notices longer wires tend to fail at lower loads
- Concludes that impurities in material contribute to failure and are more common in larger specimens
- **Does not build quantitative model, data is used to derive qualitative conclusion**





Oxygen theory of combustion



- Phlogiston theory stated that a fire-like element was contained in materials and released during combustion
- Antoine Lavoisier shows that metals increase weight undergoing combustion or reaction with the environment
- Noticed that there were "different types" of air
- Leads to the discovery of oxygen and ushers in chemical revolution

Statistics

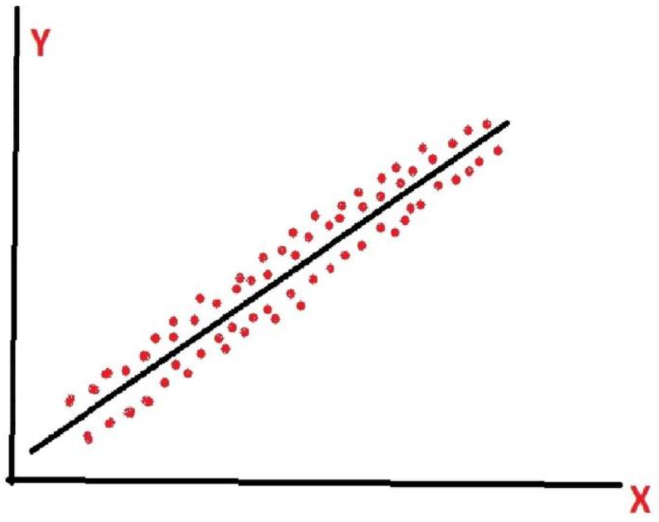
Classical statistics

- Statistical models mathematically operate on experimental data to draw quantitative conclusions
- Descriptive statistics summarizes aspects of sample, inferential statistics makes claim about the population from which data is assumed to be sampled
- **Statistical techniques often involve assuming an unknown model which explains the data, then tuning the model so that it agrees with the data**
- The model can then be used to answer further questions
- Minimal or no prior knowledge built into these analyses

Regression

- Statistical technique used to approximate functional relationship between measured input and output variables
- Typically no physical laws are known, but assumptions need to be made about the mathematical form of the relationship
- Find parameters in the regression model such that error with data is minimized
- Once model is tuned, it acts both as a "compressed" representation of the data and a predictor
- Need to assess the "goodness of fit"

Data to be fit with regression model



Assumed relationship, with error term

$$y = \alpha + \beta x + \epsilon$$

Use measurement data to write linear system that model should obey

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

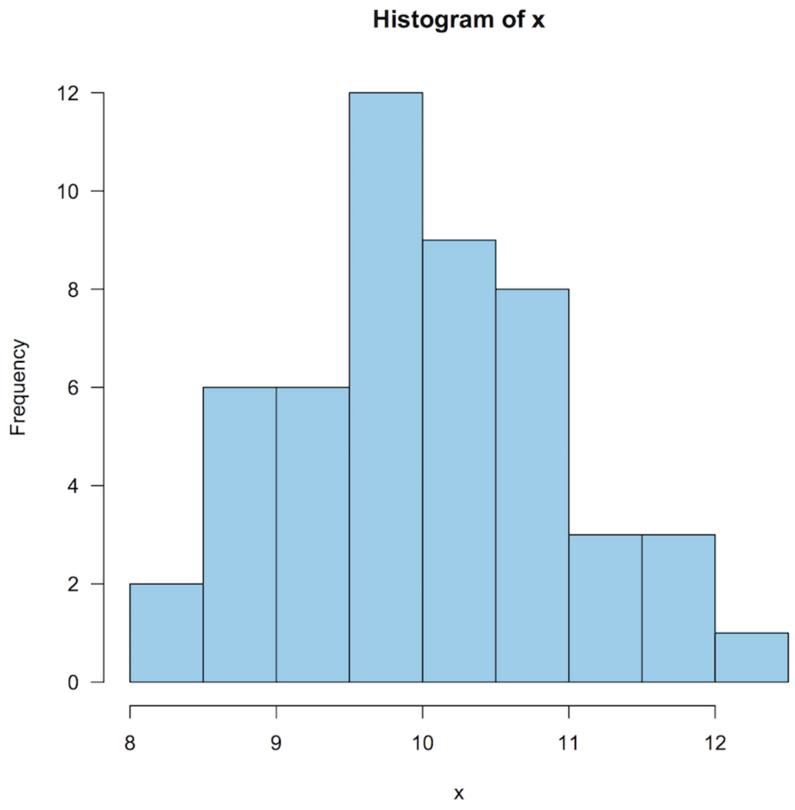
"Train" model by finding parameters that minimize magnitude of error

$$\underset{\alpha,\beta}{argmin} \left| \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \right|^2$$

Maximum Likelihood Estimation

- Want to estimate the population distribution which gave rise to a set of observations
- Find parameters of assumed population distribution
- Once a distribution is fit, it can be used to explore and/or characterize the random phenomenon
- Prior knowledge enters when we assume the form of the distribution (normal, exponential, uniform, etc.)
- Assuming samples are independent, find parameters such that the "likelihood" is maximized

Set of observations to fit with parametric distribution



$$(x_1, x_2, \ldots, x_N)$$

Assume form of population distribution

$$f = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2}$$

Construct likelihood, take its logarithm (for convenience), and optimize

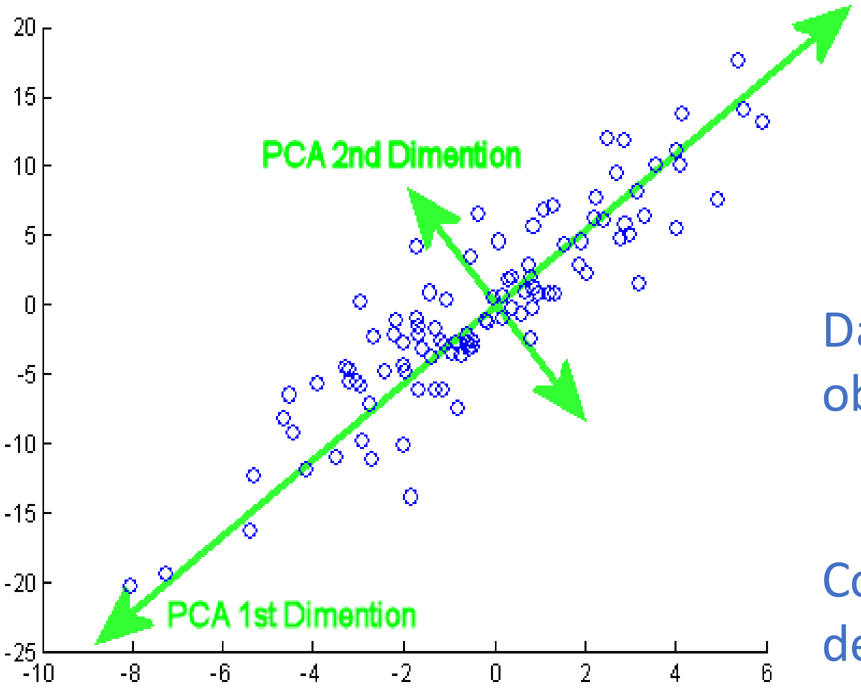
$$\mathcal{L}(x_1, x_2, \ldots, x_N; \theta) = f(x_1; \theta) * f(x_2; \theta) * \cdots * f(x_N; \theta)$$

$$\log \mathcal{L} = \sum_i \log\big(f(x_i; \theta)\big)$$

$$\underset{\theta}{argmin} \sum_i - \log\big(f(x_i; \theta)\big)$$

Dimensionality Reduction

- High-dimensional data sets can lie in low-dimensional subspaces (exactly or approximately)
- Principal component analysis is common approach to find a basis for the low-dimensional subspace
- These efficient representations can be used to gain insight, for data storage, denoising, etc.
- Find directions of maximum variance in the data to compute "principal components"



Data with low-dimensional representation

Data matrix with N observations

$$\underline{\underline{X}} = \begin{bmatrix} - & \underline{x}_1 & - \\ & \vdots & \\ - & \underline{x}_N & - \end{bmatrix}$$

Coordinates in basis defined by first principal component

$$\underline{t}_1 = \begin{bmatrix} - & \underline{x}_1 & - \\ & \vdots & \\ - & \underline{x}_N & - \end{bmatrix} \begin{bmatrix} | \\ \underline{w}_1 \\ | \end{bmatrix}$$

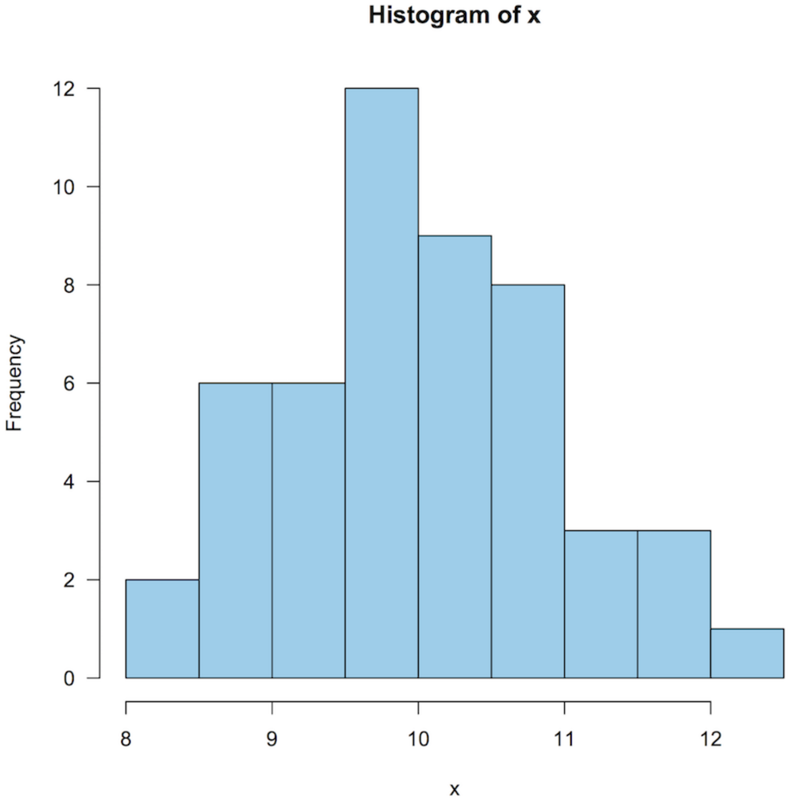
Direction of first principal component found by maximizing variance

$$\underset{||w_1||=1}{argmax}\left(\underline{t}_1 \cdot \underline{t}_1\right) = \underset{||w_1||=1}{argmax}\left(\underline{\underline{X}}\,\underline{w}_1 \cdot \underline{\underline{X}}\,\underline{w}_1\right)$$

Hypothesis Testing

- Common statistical method used to determine the probability that observed data came from a given distribution
- Can also test whether different data sets came from the same distribution
- Used frequently in medicine and psychology to see if a treatment had an effect which was unlikely to be from chance (statistically significant)
- Fits distribution to data and then uses that fit to draw conclusions
- Consider t-test: we want to find probability that the mean of an observed sample came from a population with given mean and variance

Does this data come from a distribution with mean μ?



$$(x_1, x_2, \ldots, x_N)$$

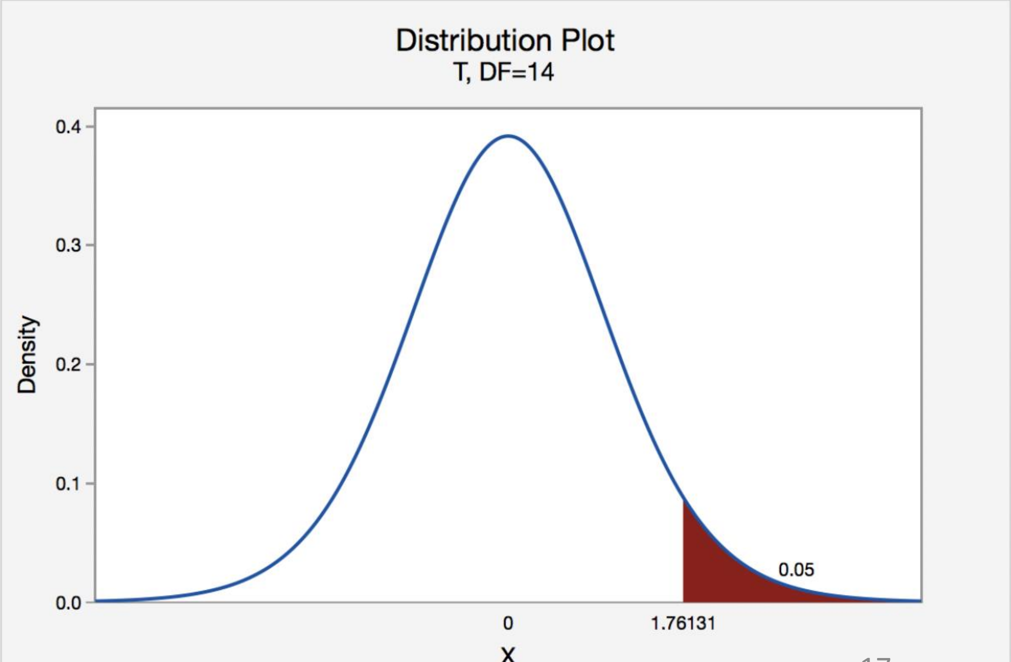
Null and alternative hypotheses

$$H_0 : E(X) = \mu, \quad H_A : E(X) > \mu$$

Test statistic discounting difference between sample mean and population mean with variance

$$t = \frac{\frac{1}{n}\sum_n x_n - \mu}{\sigma/\sqrt{N}}$$

Null hypothesis rejected when test statistic exceeds threshold



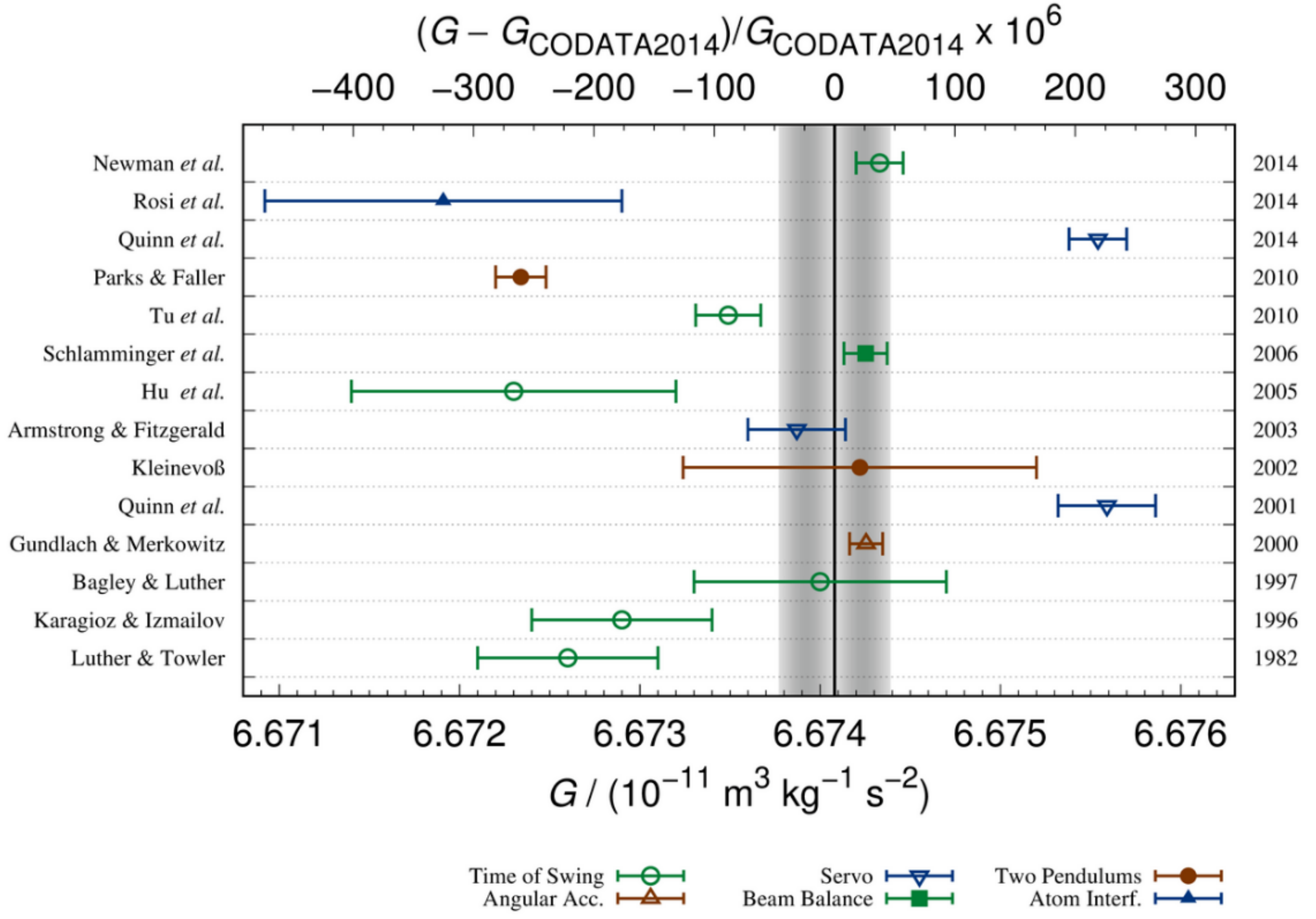
Physics

Classical physics

- Algebraic and differential equations relate quantities of interest
- Data is used indirectly to construct the general form of these equations
- Empirical parameters always show up, and are fit to experimental data directly
- Have to ask the question: if the model were correct, what would the empirical parameter be?
- We "train" the model by calibrating empirical parameters to data, then "test" it by seeing how well it generalizes

Newtonian Gravity

- Gravitational attraction is proportional to square of distance and the two masses of bodies, but scaled by an empirical factor
- This factor needs to be determined by data through experiments
- Calibrate on dynamics of one configuration of masses, test in novel application and ensure model is predictive

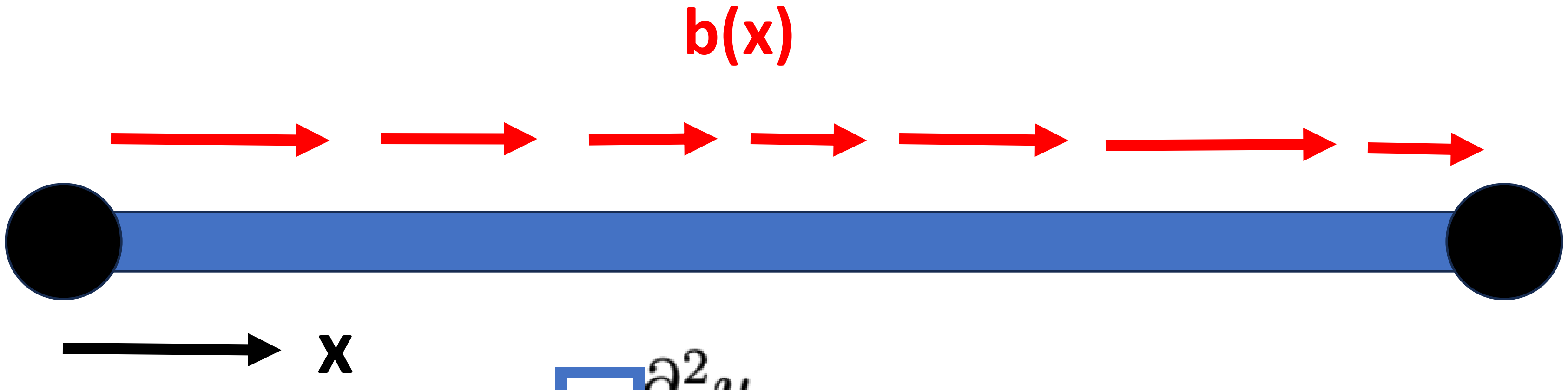


$$\underline{F}(\underline{r}) = G\frac{m_1 m_2}{\underline{r} \cdot \underline{r}}\hat{\underline{r}}$$

Constitutive Models in Continuum Mechanics

- Relationships between stress and strain measures "close" system of equations which describe mechanical equilibrium in solids and fluids
- Represent properties of material which cannot be determined from theoretical considerations alone
- Need to experimentally determine constitutive relations
- Have to make assumptions about the form of the material response (linear, non-linear, viscoelastic, plastic, etc.)

Finding constant stiffness of elastic bar



Governing equation with unknown stiffness

$$\boxed{E}\frac{\partial^2 u}{\partial x^2} = b(x)$$

Solution to governing equation

$$\implies u(x, E) = \frac{1}{E}\int_0^x \int_0^y b(z)dzdy$$

Measured displacement data at given position

Loss function

$$\underset{E}{argmin}\sum_i \left(\frac{1}{E}\int_0^{x_i}\int_0^y b(z)dzdy - \boxed{\hat{u}(x_i)}\right)^2$$

Finding spatially varying stiffness of elastic bar

Governing equation with unknown stiffness

$$\frac{\partial}{\partial x}\left(E(x;\theta)\frac{\partial u}{\partial x}\right) = b(x)$$

Stiffness now varies in space and is parameterized

$$\boxed{E(x,\theta) = \sum_{i=0}^{N} \theta_i x^i}$$

Solution to governing equation

$$\implies u(x,\theta) = \int_0^x \frac{1}{E(y;\theta)} \int_0^y b(z)dzdy$$

Displacement data at given position

Loss function

$$\underset{\underline{\theta}}{argmin} \sum_i \left(\int_0^{x_i} \frac{1}{E(y;\underline{\theta})} \int_0^y b(z)dzdy - \boxed{\hat{u}(x_i)} \right)^2$$

Data-driven modeling

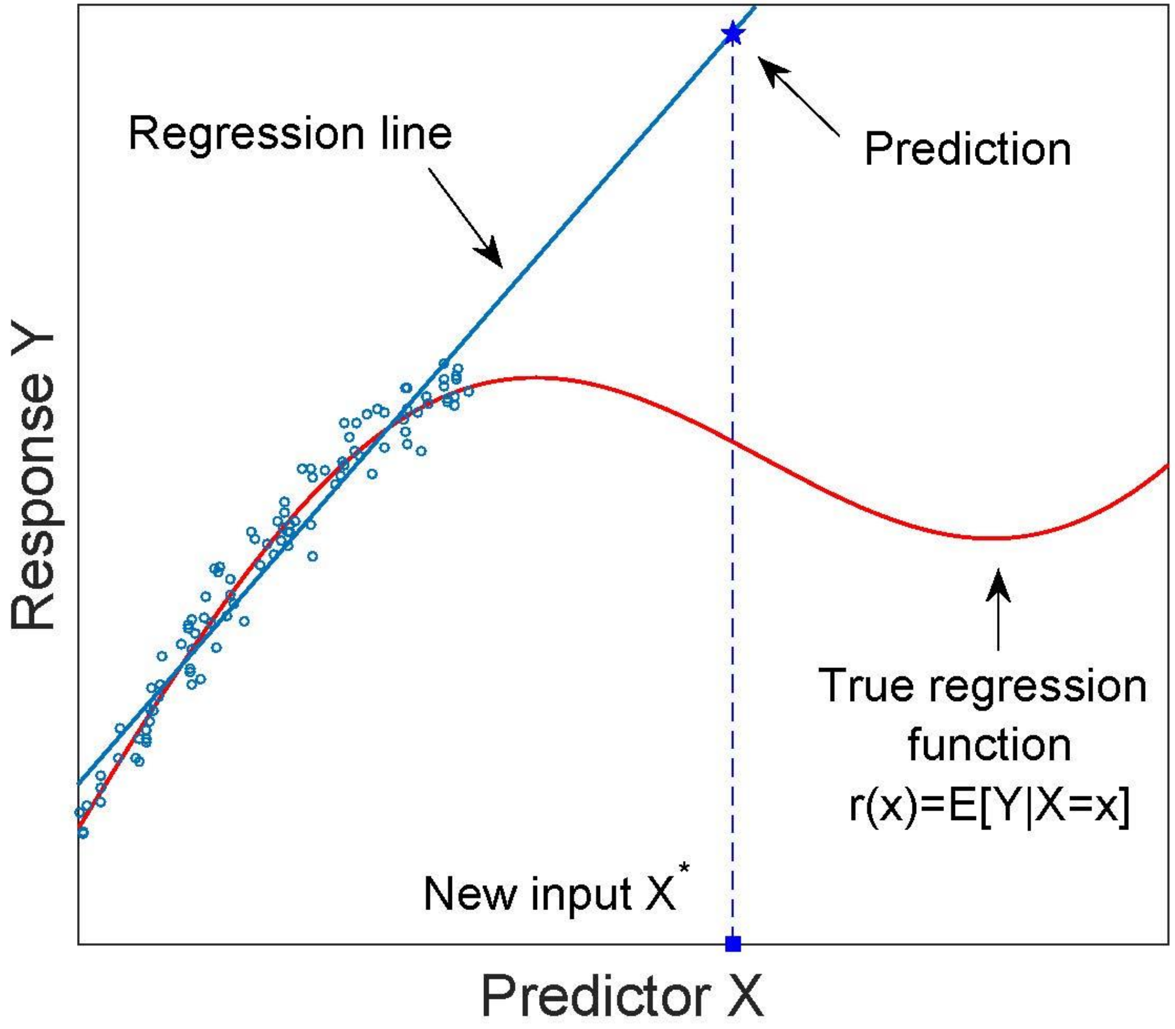
Modern data-driven modeling

- Collection of diverse techniques asking the question ***"how do we build/inform physical models with data?***"
- Anywhere from using data to discovering physical laws, to streamlining numerical solutions, to fully replacing numerical solutions
- Not all machine learning—machine learning typically refers to the explicit use of neural networks, not simply making use of data
- Exciting and productive area of research because data is more available, hardware and software infrastructure is relatively new, and traditional modeling and solution techniques have been heavily researched for decades
- Data-driven models make prediction possible even when underlying laws are not known

Surrogate Models

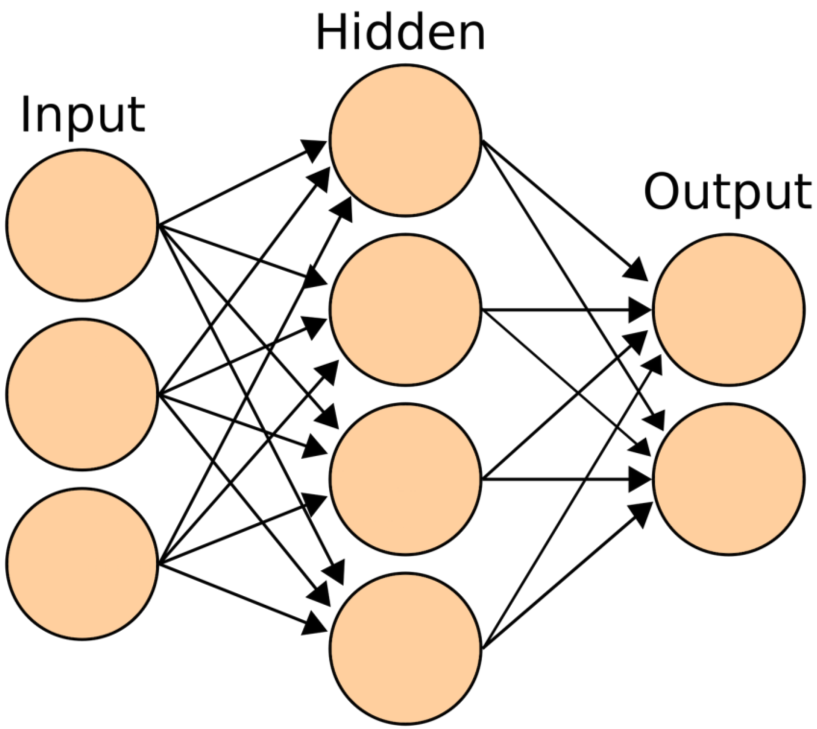
- What people typically think of when hearing "machine learning"
- Black-box model with large number of parameters trained on big data set to find input-output relationships
- Neural networks are very general non-linear function approximators
- Python libraries make building and training these models easy
- No prior knowledge or constraints applied to model other than trying to fit training data set
- Can think of these models as good at "interpolating" but not necessarily reliable when "extrapolating"

Interpolation vs. Extrapolation



Want to ensure that model is used to interpolate, otherwise there is no guarantee the fit is trustworthy. But, it may be difficult to precisely distinguish between interpolation and extrapolation.

Multilayer perceptron neural network



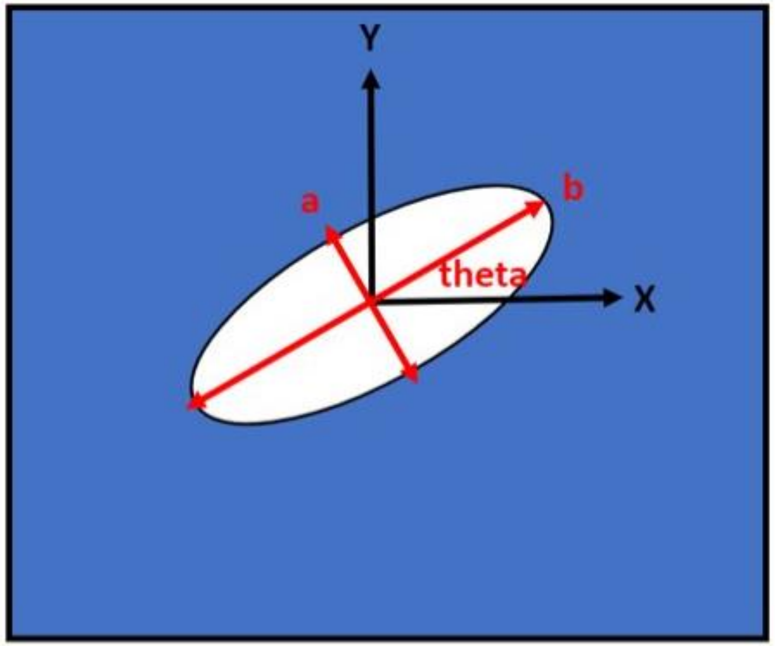
Output "O" computed with series of linear transformations on input "I," with non-linear "activations" applied at each hidden layer

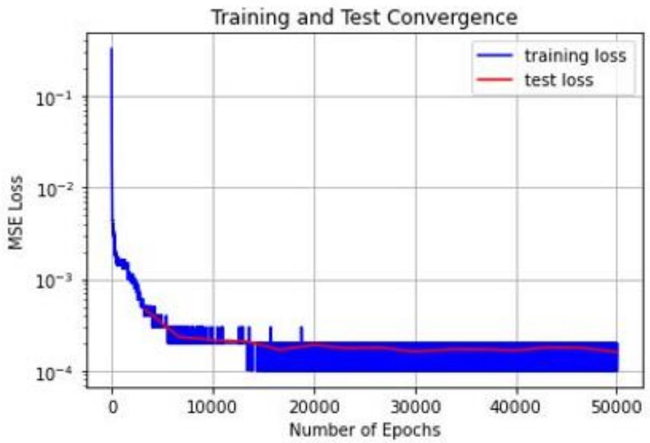
$$O_i = w_{ij}^2 \sigma(w_{jk}^1 I_k + b_j^1)$$

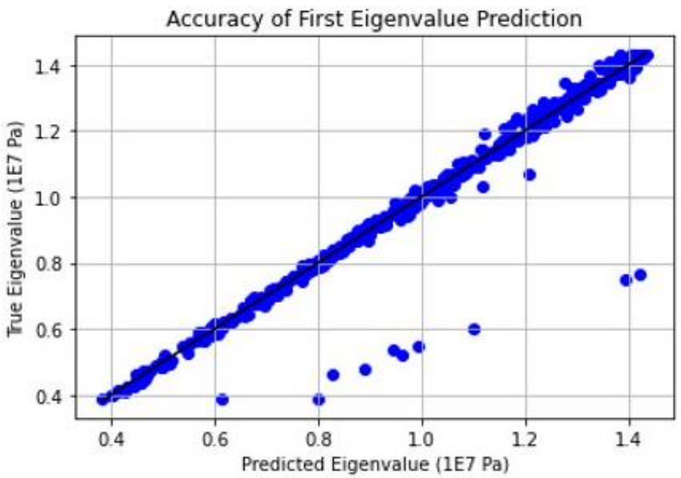
$$\underset{\underline{\underline{w}}^2, \underline{\underline{w}}^1, \underline{b}^1}{argmin} \sum_i \left| \underline{\underline{w}}^2 \sigma(\underline{\underline{w}}^1 \underline{I}^i + \underline{b}^1) - \underline{\hat{O}}(\underline{I}^i) \right|^2$$

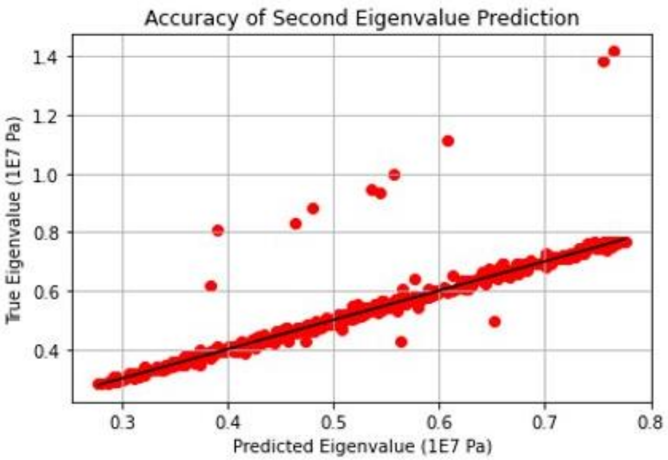
Parameters of network (weights and biases) are determined by minimizing the difference between the data and the predictions of the network

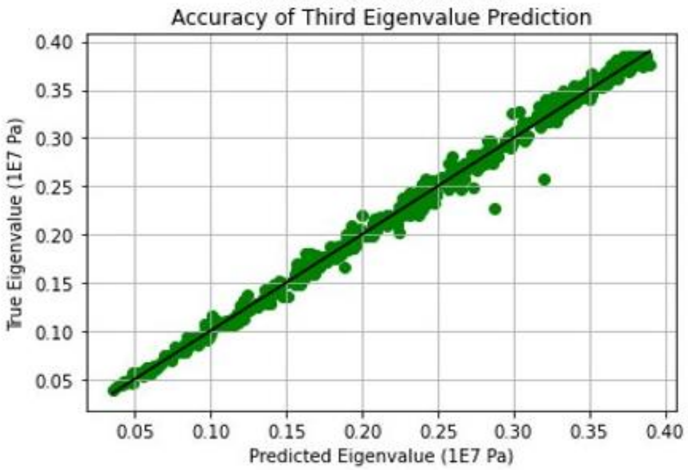
Example: Effective Material Properties











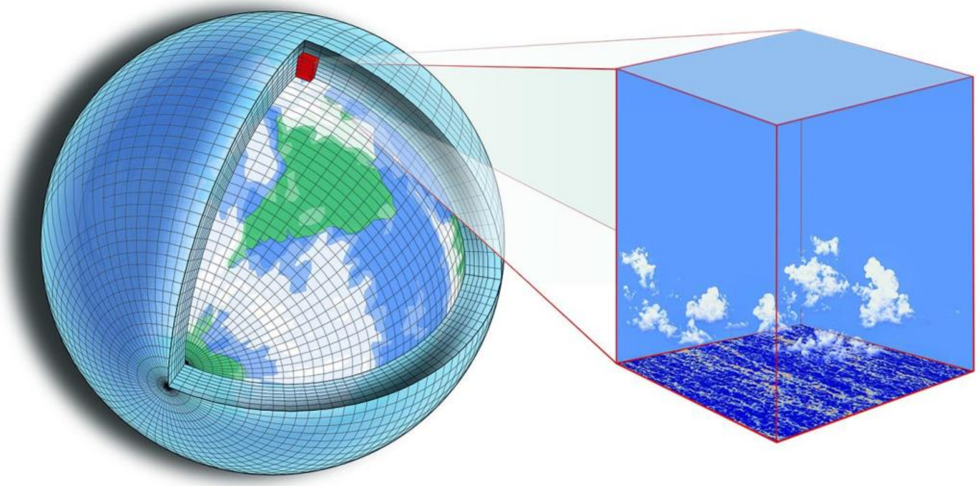
$$\begin{bmatrix} a \\ b \\ \theta \end{bmatrix} - \text{neural network} \rightarrow \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}$$

Train surrogate model to map relation from material microstructure to effective (homogenized) material properties

Parameter estimation

- Neural networks are used to parameterize and fit functions in physical models
- Neural networks simply act as flexible ways to represent diverse functional behavior; not a novel idea
- Could be learning aspects of model which are empirical anyway (constitutive relations)
- The point is to learn some aspect of a physical model, not replace the model
- Applications include: hyperelasticity, viscoelasticity, damage models, climate models

Computational modeling of climate influenced by sub-grid features like clouds



Add unknown forcing term to incompressible Navier-Stokes equations to account for sub-grid features

$$\frac{\partial u_i}{\partial t} = -u_j \frac{\partial u_i}{\partial x_j} - \frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} + \mathcal{N}_i(p, u_1, u_2; \theta)$$

$$\frac{\partial u_j}{\partial x_j} = 0$$

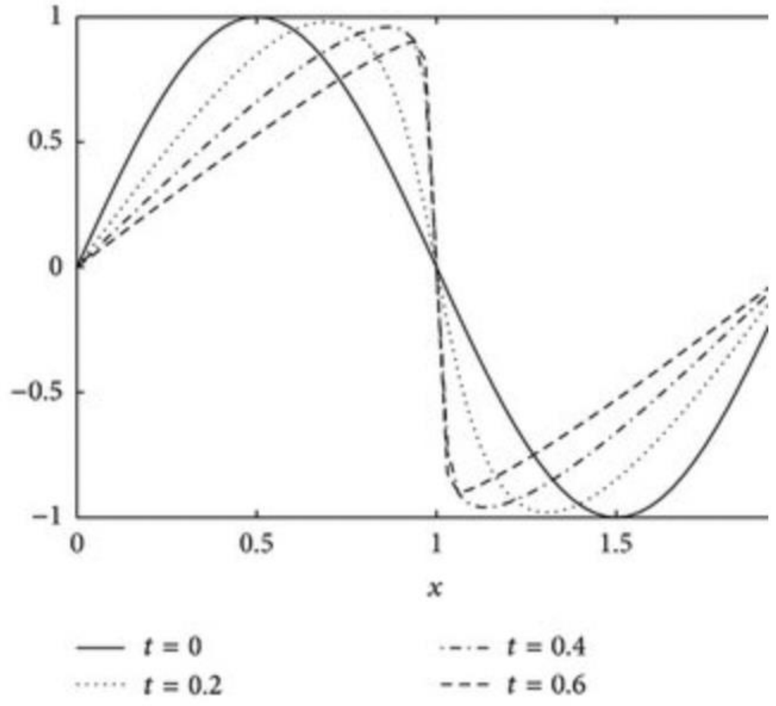
Determine forcing term by fitting to climate data set..."S" is short-hand for solution to Navier-Stokes

$$\underset{\theta}{argmin} \sum_i \sum_j \sum_k \sum_\ell \left(\hat{u}_i(x_j, y_k, t_\ell) - S_i(x_j, y_k, t_\ell; \theta) \right)^2$$

Equation Discovery

- Use data from a dynamical system to learn a governing partial differential equation
- Typically assume the order of time derivative and that there are no mixed space/time derivatives
- Represent the spatial "forcing" part of the equation as a neural network and use measurement data to estimate it
- It is convenient to introduce an intermediate step of fitting another neural network to the data itself, which makes differentiation easy

Want to discover differential equation that describes observed space/time dynamics (Burger's equation)



Represent RHS of PDE with neural network, and fit another neural network to the data itself

$$\frac{\partial u(x,t)}{\partial t} = \mathcal{N}\left(x, t, u, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}; \theta\right)$$

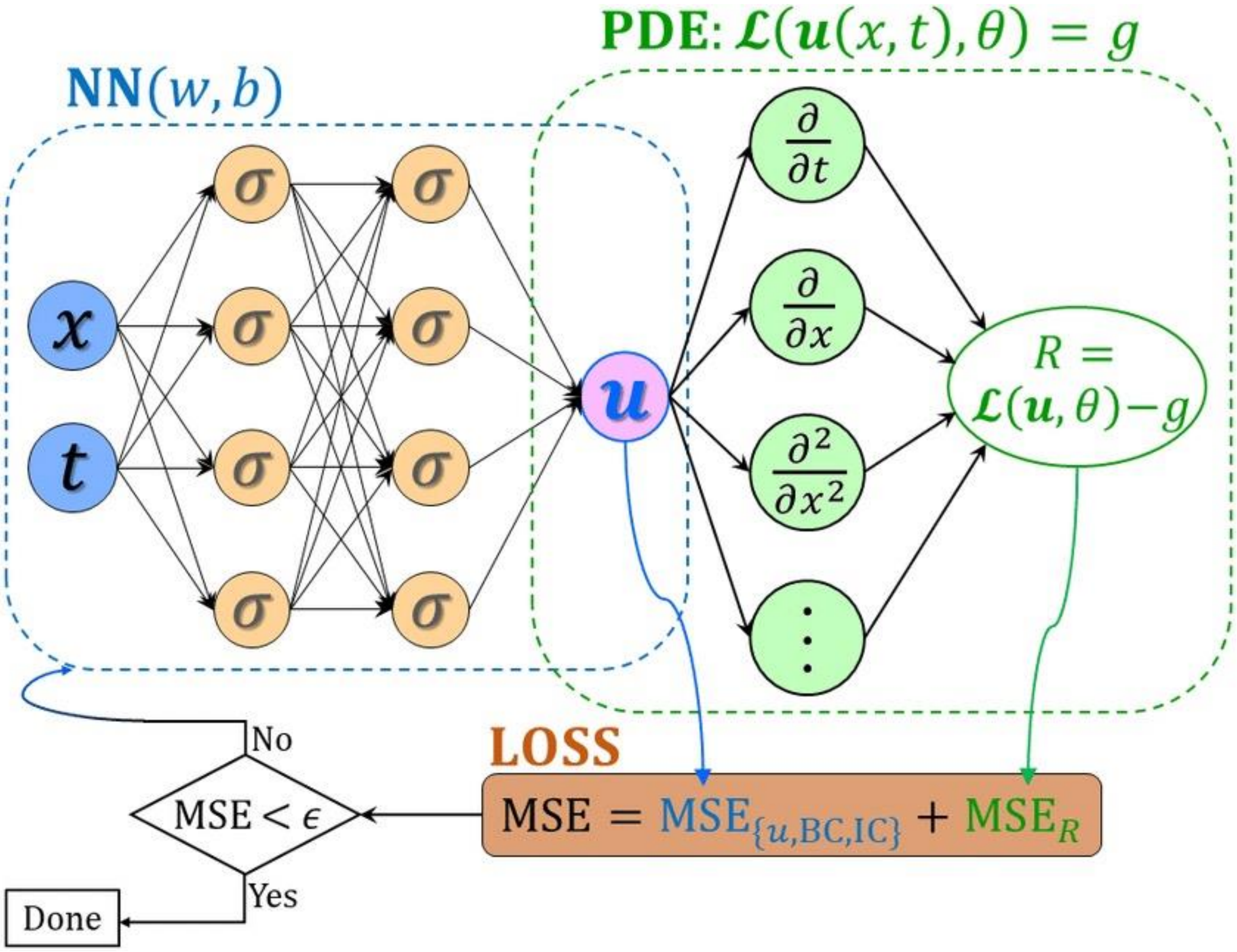
$$u(x,t) = \mathcal{U}(x, t; \Phi)$$

Simultaneously minimize loss between the data and its fit and the time derivative of the fit with RHS network

$$\underset{\Phi,\theta}{argmin}\left[\lambda_1 \sum_i \sum_j \left(\hat{u}(x_i,t_j) - \mathcal{U}(x_i,t_j;\Phi)\right)^2 + \lambda_2 \sum_i \sum_j \left(\frac{\partial \mathcal{U}(x_i,t_j;\Phi)}{\partial t} - \mathcal{N}(x_i,t_j,u(x_i,t_j),\ldots;\theta)\right)^2\right]$$

Data Assimilation

- Data assimilation is a term for integrating measurement data with governing equations of a system
- Originally comes from weather models, where sparsely measured initial conditions often lead to unstable solutions
- Physics-informed neural networks (PINN's) are one technique for incorporating knowledge of the underlying physics into a data-driven model
- This is useful when data is sparse—a surrogate model with many parameters will overfit small training data, but including physical constraints helps regularize the optimization problem
- Network should match training data AND satisfy governing equations at select "collocation" points



The total loss penalizes discrepancies with a sparse training data set, and failures of the network to satisfy boundary/initial conditions and the governing equations. Parameters of the network are computed by minimizing this loss function.

Model Reduction

- Model reduction uses data to reduce the computational cost of numerical simulations of PDE's
- Principal Orthogonal Decomposition (POD) is a common technique which computes a set of global spatial shape functions from measurement data on a dynamical system
- For some systems, it is possible to find a small number of shape functions which accurately capture the dynamics
- Using a small set of global shape functions as a discretization can greatly speed up simulations
- We still solve the same governing equations!
- Have to assume that basis computed from one set of boundary data and forcing is still a good choice in different contexts

Data matrix contains "snapshots" of solution in space in each row

$$\underline{\underline{X}} = \begin{bmatrix} u(x_1,t_1) & u(x_2,t_1) & \dots & u(x_N,t_1) \\ u(x_1,t_2) & u(x_2,t_2) & \dots & u(x_N,t_2) \\ \vdots & \vdots & \dots & \vdots \\ u(x_1,t_T) & u(x_2,t_T) & \dots & u(x_N,t_T) \end{bmatrix}$$

...want to find spatial vector(s) which explain the most variance in data matrix...

Looks very similar to PCA dimensionality reduction problem!

$$\underline{\underline{X}} = \begin{bmatrix} - & \underline{x}_1 & - \\ & \vdots & \\ - & \underline{x}_N & - \end{bmatrix}$$

$$\underline{t}_1 = \begin{bmatrix} - & \underline{x}_1 & - \\ & \vdots & \\ - & \underline{x}_N & - \end{bmatrix} \begin{bmatrix} | \\ \underline{w}_1 \\ | \end{bmatrix}$$

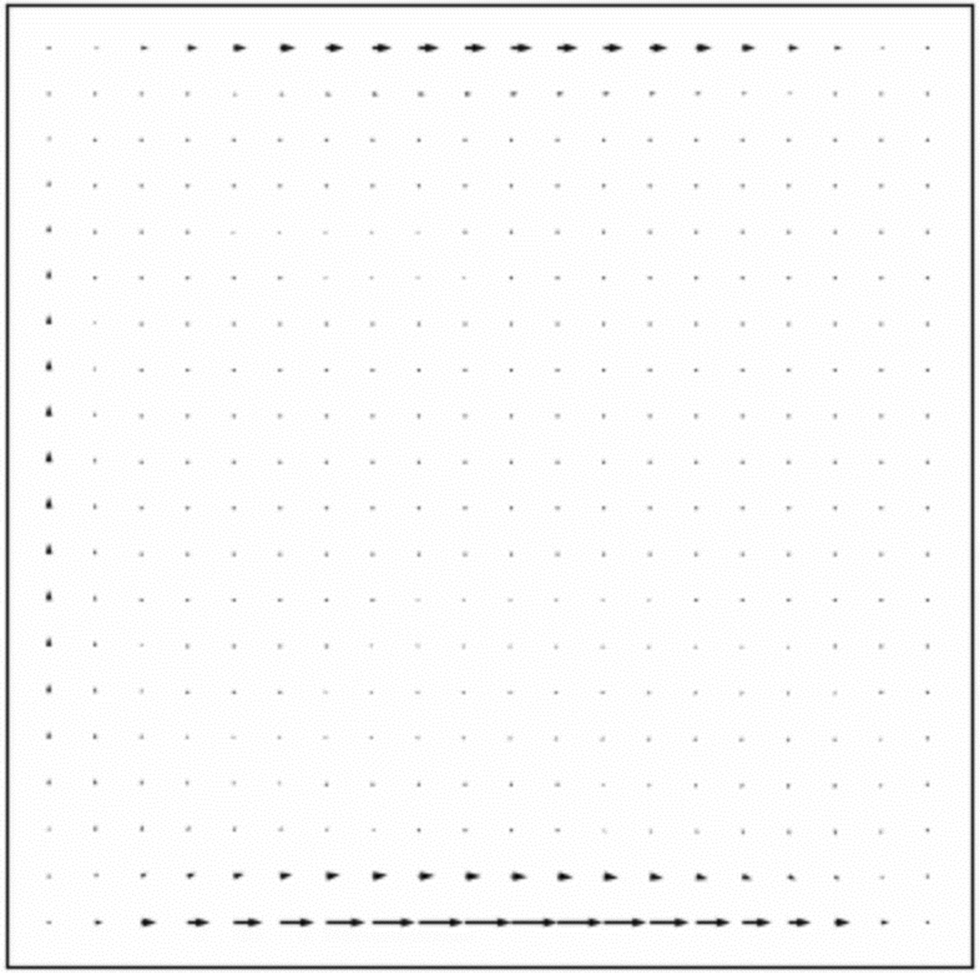
$$\underset{||w_1||=1}{argmax}\left(\underline{t}_1 \cdot \underline{t}_1\right) = \underset{||w_1||=1}{argmax}\left(\underline{\underline{X}}\underline{w}_1 \cdot \underline{\underline{X}}\underline{w}_1\right)$$

Solution then discretized with "principal components"

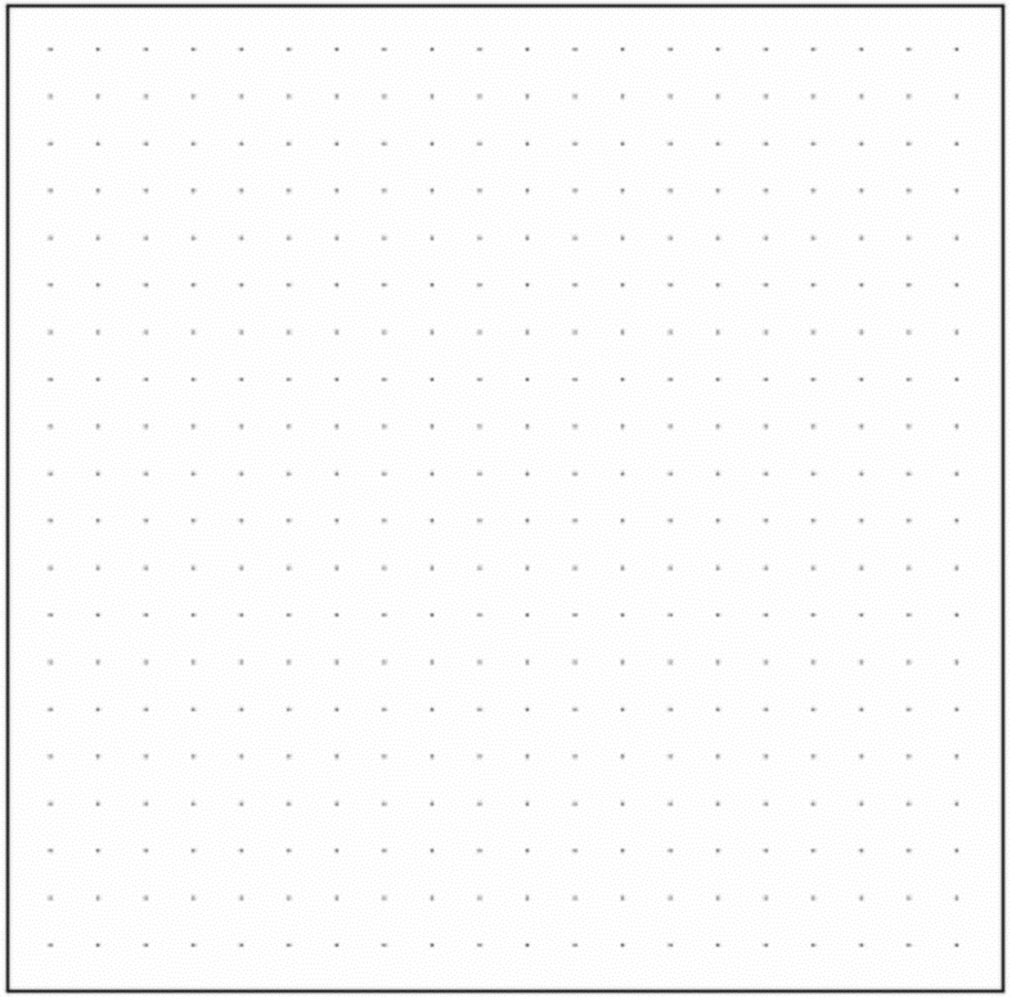
$$\underline{u}(t) = \sum_n a_n(t)\underline{w}_n$$

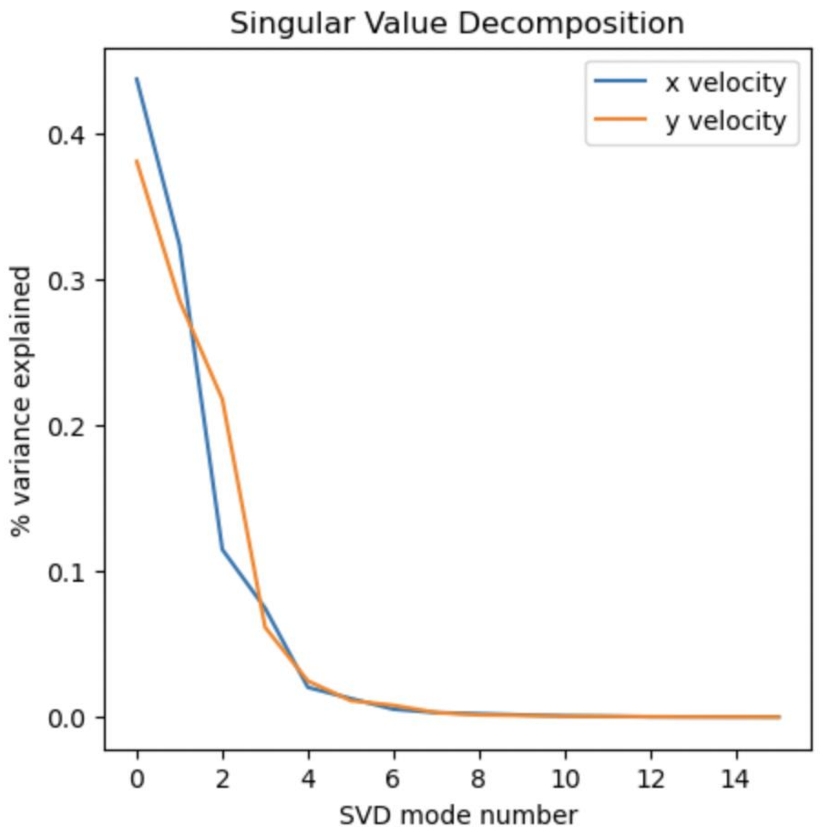
Comparison of flow solution to reduced model

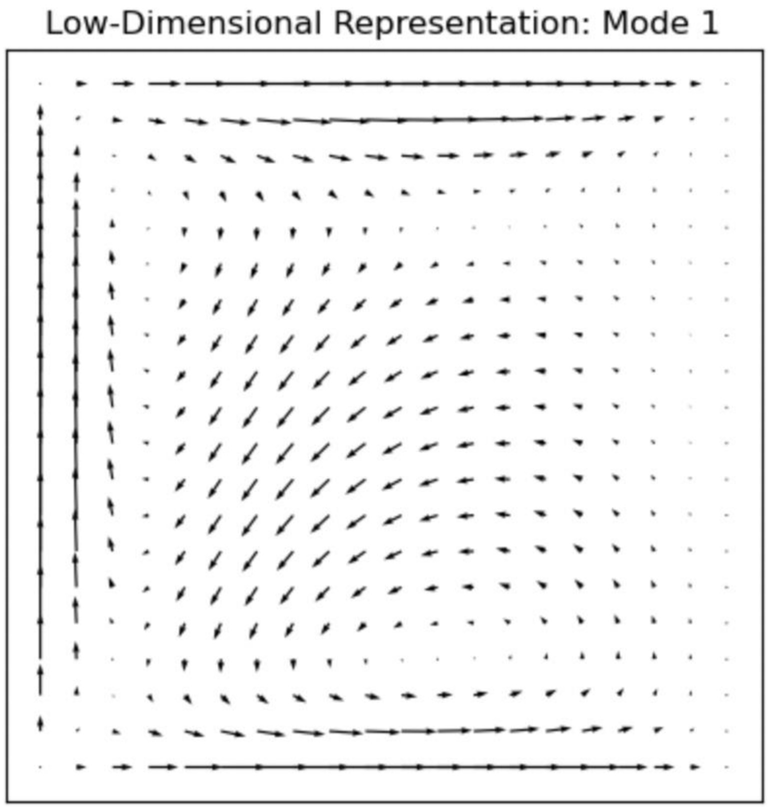
True solution

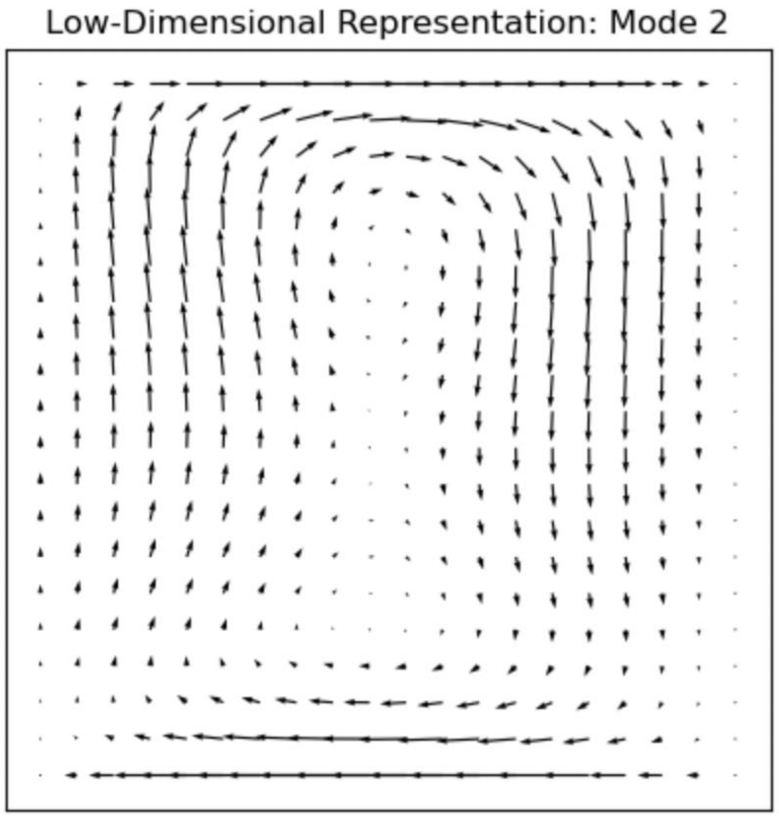


Solution reconstructed with linear combination of 3 POD modes









Singular values ordered least to greatest, fast decay indicates low-dimensional strucutre

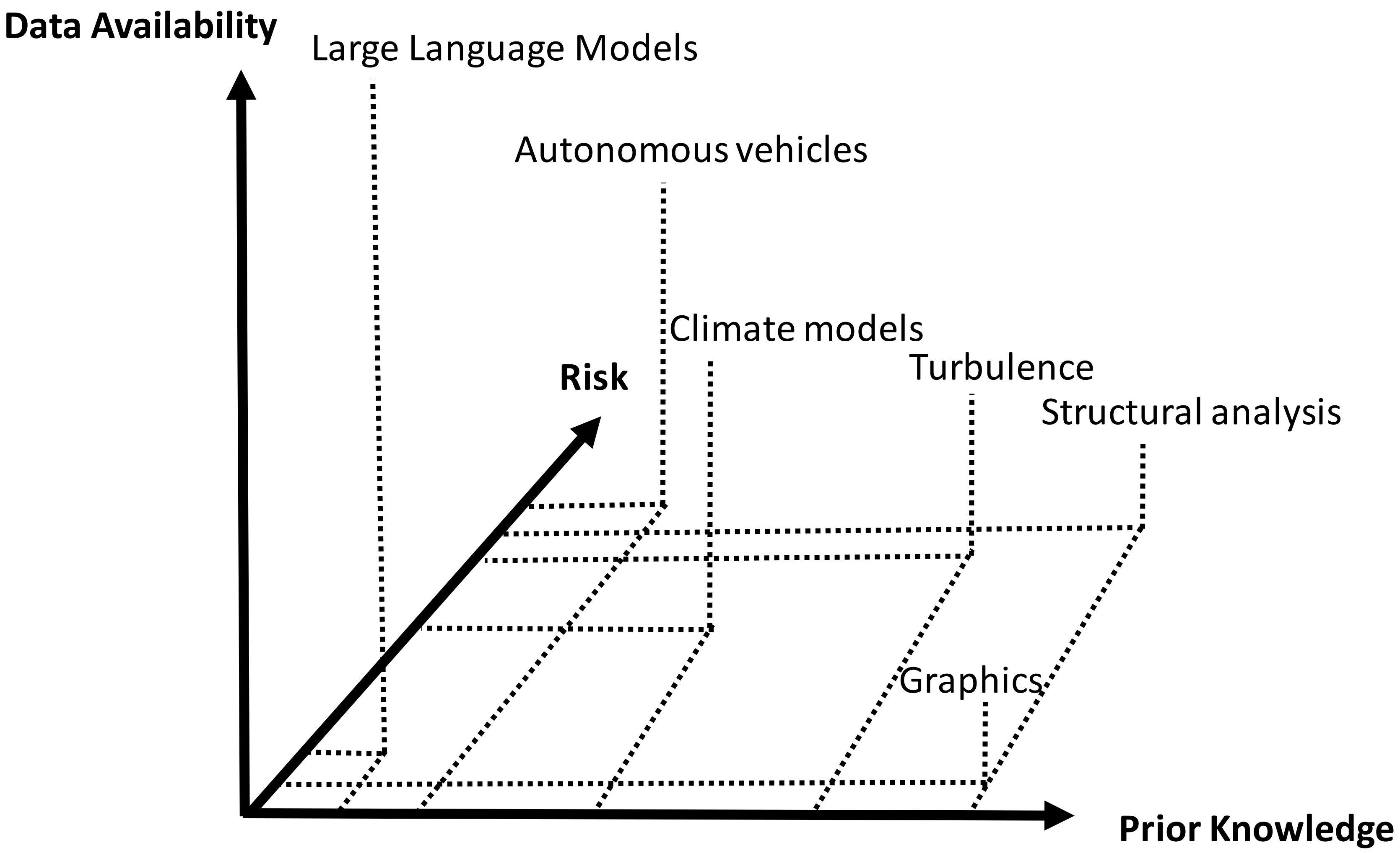
First two POD spatial modes

Categorizing Applications

Assessing the Model

- Less prior knowledge -> less structure -> more free parameters -> more data required to train -> more concern about generalization
- **Different application areas have different quantities of available data, prior knowledge, and expectations for model performance**
- Some applications have zero tolerance for error, while others are more lenient
- The possibility and appropriateness of a model (i.e. machine learning, traditional physics, or some combination) is highly context-dependent
- Useful to have some way of categorizing the context in which a model is used

Categorizing common models



- Could additional dimension of "speed requirements" to categorization (online vs. offline use)
- Physical models excel at any risk level, high prior knowledge, and low data availability
- Data-driven techniques tend to excel with low risk, low prior knowledge, and high data availability
- When data is abundant, surrogate models can perform well with no prior knowledge input
- ML still not being used in situations with high-risk
- Tricky application when there is some prior knowledge and some data available

Data-physics spectrum

Black-box data-driven surrogate models

Not always clear what lives here, hard to continuously interpolate between two extremes!

Physics and traditional numer-ical methods

Model Type

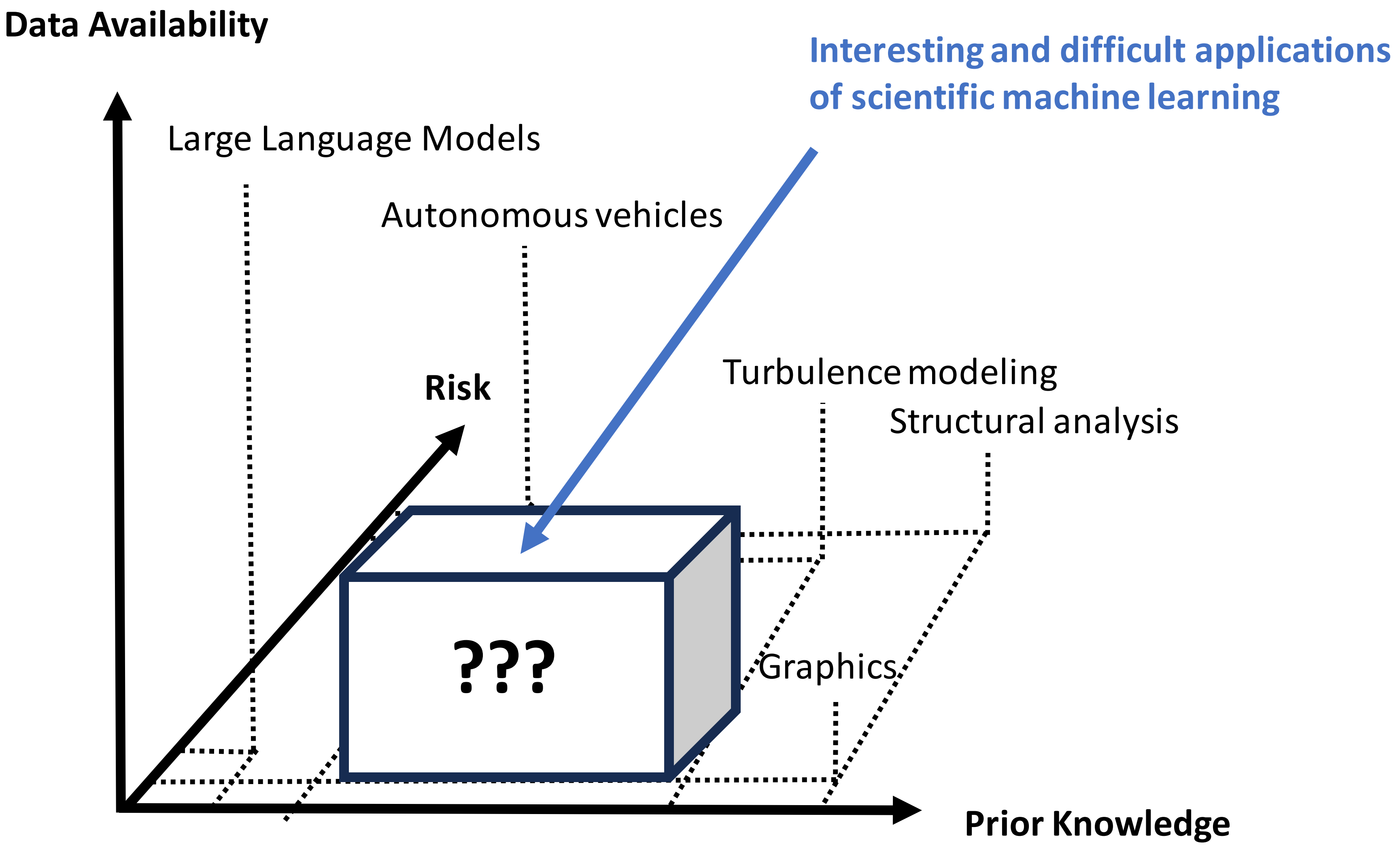
Features

Large data sets, flexible models, large number of empirical parameters, no prior knowledge

Some data, but models need to be supplemented and constrained by prior knowledge

Sparse data, highly structured models, small number of empirical parameters, lots of prior knowledge

Where are opportunities?



Ideas for incorporating ML and physics

- **Data-driven accelerator:** initializing design in topology optimization, newton solve initialization, hyperparameter optimization
- Prediction of surrogate model always corrected by physics, but still speeds up iterative solution process
- **Constitutive modeling:** replace empirical aspects of models with flexible function approximators and learn them from data
- **PINN's with relaxed constraints:** how to enforce governing equations whose form you don't precisely know?

Takeaways

- Data very frequently used to tune parameters in a model by solving an optimization problem
- Physics models and machine learning differ in how much prior knowledge and structure go into the model; both are "trained" and "tested" on data
- In general, we might think that linear models with fewer parameters lead simultaneously to less data requirement and more reliable generalization
- Lots of machine learning work does not seem to consider "risk" as influential in whether the model will be used
- Important to understand the nature of the application area and choose an approach accordingly
- Whether a certain model type is possible and/or useful is very situational
- **Not a bright line between data-driven and physics-based models!**