Show Me A Function: More Than Meets The Eye

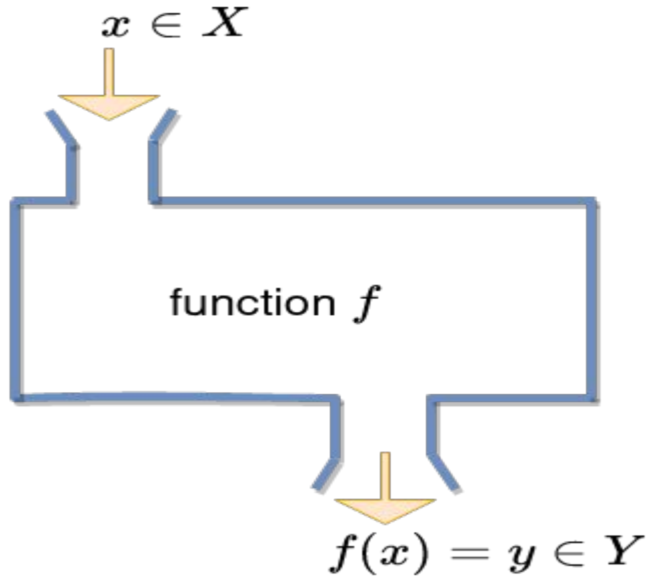
by

Chinedu Eleh

(Advisor: Dr. Hans Werner van Wyk)

Definition of Function

➢ A function from a set X to a set Y assigns to each element of X exactly one element of Y



Using Simple Functions to Build Complex Ones

- Functions we learn in precalculus, calculus, etc
  - polynomials
  - exponential
  - trigonometric
  - inverse
  - composite functions, etc

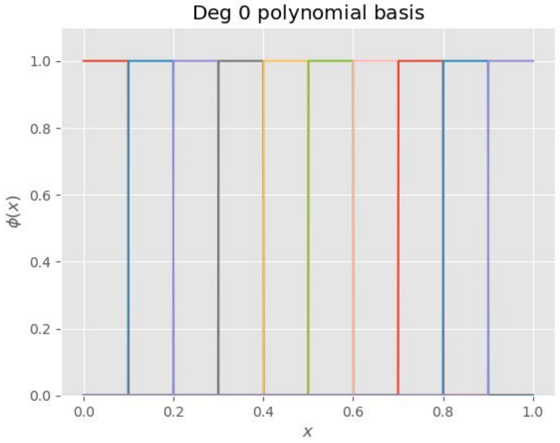
Polynomial Functions

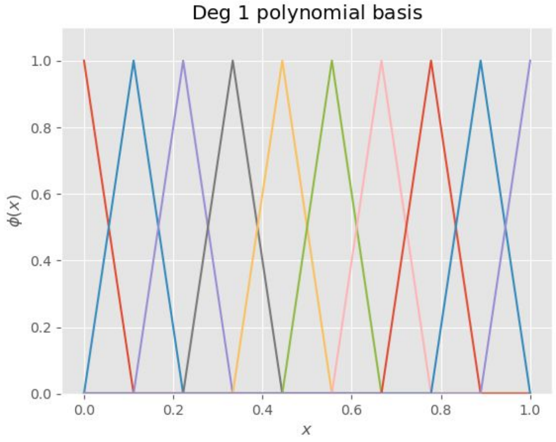
- ➢ Spline Interpolation
- ➢ Finite elements
- ➢ ReLU activation function

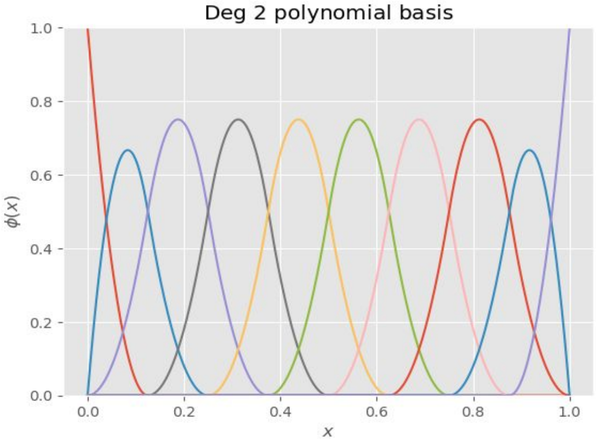
Spline Basis

Let ϕ_i be the indicator function of $[x_i, x_{i+1}]$.

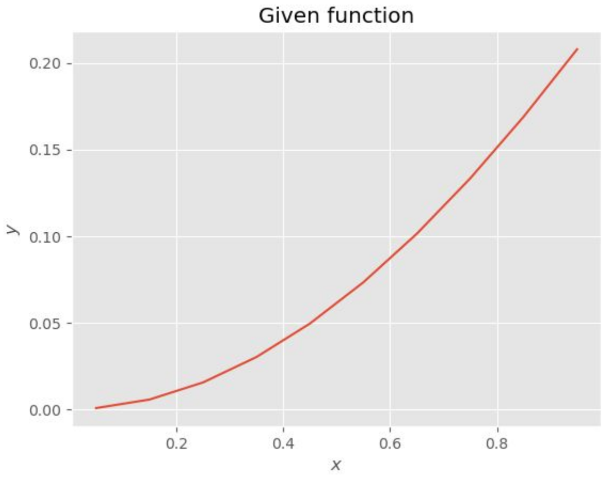
$$\phi_i(x) = \begin{cases} 1, & x \in [x_i, x_{i+1}] \\ 0, & \text{otherwise} \end{cases}$$

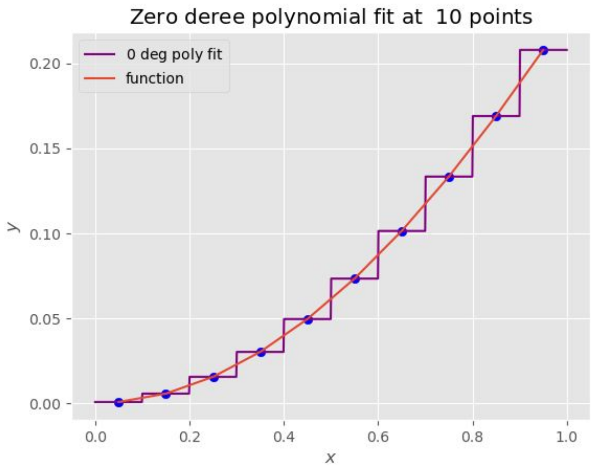


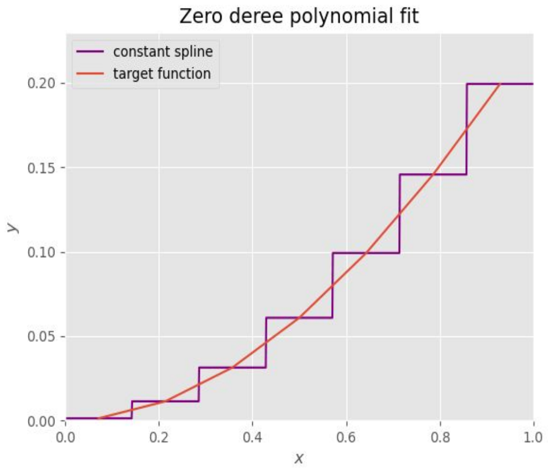




Spline Interpolation



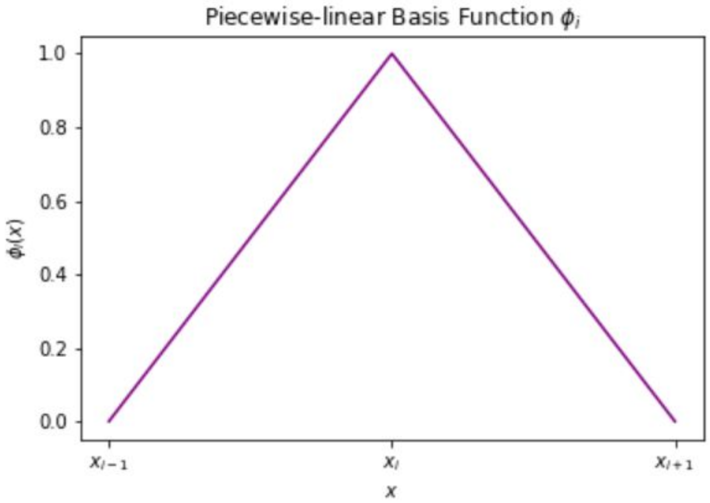




Finite Element Basis

- ➢ Finite elements method built on similar idea of spline basis
- ➢ Basis could be constant, linear, quadratic, or higher order piecewise poly

$$\phi_i(x) = \begin{cases} 0, & x < x_{i-1}; \\ \frac{x - x_{i-1}}{h} & x_{i-1} \leq x < x_i; \\ \frac{x_{i+1} - x}{h} & x_i \leq x < x_{i+1}; \\ 0, & x_{i+1} \leq x. \end{cases}$$



Finite Element Approximation

- Goal is to solve boundary value problems, say

$$-\nabla \cdot (q(x)\nabla u) = f(x), \;\; x \in \Omega,$$
$$u = 0, \;\; x \in \partial\Omega$$

- Discretization of the form

$$u(x) = \sum_{i=i}^{n} c_i \phi_i(x), \;\; \phi_i(x) \in V^h$$

is assumed, where V^h, spanned by ϕ_i, is a finite dimensional approximation of the unknown infinite dimensional space.

Differential Form ⇨ Weak Form ⇨ Discretization ⇨ Linear System

ReLU Activation

➢ The ReLU activation function is defined as

$$\phi(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

➢ It can create sufficient nonlinearities in neural network layers to learn virtually any mapping

Neural Networks

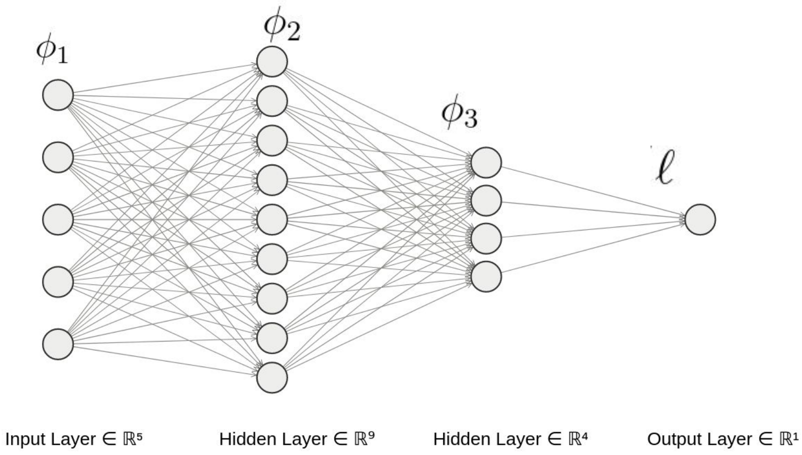
- Made up of composition of affine functions with activations creating nonlinearity where necessary

$$f(x) = (\ell \circ \phi_3 \circ \phi_2 \circ \phi_1)(x)$$

$$\phi_i(x) = \sigma_i(W_i x + b_i)$$

- ❑ σ_i is one of sigmoid, tanh, ReLU, linear, etc functions
- ❑ ℓ is mostly linear, sigmoid, softmax depending if a regression, binary classification, or multiclass classification problem

- Can take any tensor input (CNN, RNN, VAE, etc)



Trigonometric Functions

- Very well applicable in
  - Fourier transform
  - Activation functions (sinc)
  - Anywhere periodicity is desired

Exponential Functions

- Exponential growth and decay
- Density
- Kernels (SVM, RKHS, covariance)
- Activation functions (sigmoid, softmax)
- Wavelets

Inverse Functions

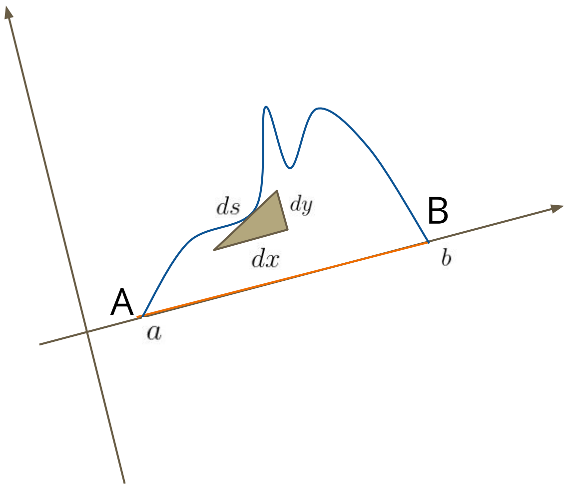
- Activation functions (arctan)
- Loss function (e.g. log in cross entropy, KL divergence)

Function Discovery

- Three main ways of discovering new functions
  - Calculus of variations
  - Statistics
  - Differential equations
- Calculus of variations date back to Euler and Lagrange, statistical methods and differential equations have rich history as well, but part of state of the art

Calculus of Variations

➢ The Classical Isoperimetric Problem: Determine a curve with a length of s that connects points A to B, such that when combined with the line segment AB, forms the largest possible enclosed area.



Length of curve:

$$ds^2 = dx^2 + dy^2$$

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

$$s = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

Area to be maximized:

$$A = \int_a^b y(x)dx$$

$$\underset{y \in \mathcal{F}}{\text{maximize}} \int_a^b y(x)dx$$
$$\text{s.t} \int_a^b \sqrt{1 + (y')^2}\, dx = s$$

Euler-Lagrange Equations

- The maximizer of the constrained optimization is a section of the circle

$$(x - c_1)^2 + (y - c_2)^2 = r^2$$

which is a solution to the Euler-Lagrange (differential) equation

$$\boxed{\frac{\partial}{\partial y}\mathcal{L}(x, y, y') - \frac{d}{dx}\left(\frac{\partial}{\partial y'}\mathcal{L}(x, y, y')\right) = 0}$$

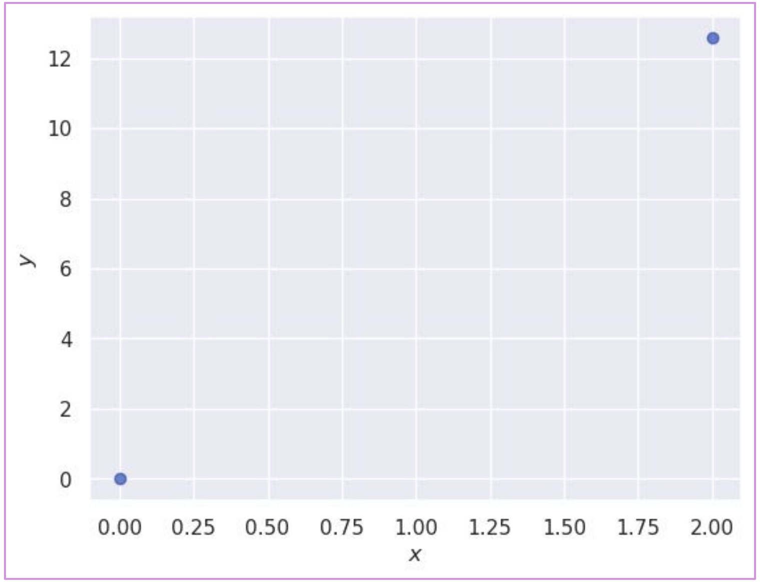
- The isoperimetric problem is solved by a function.

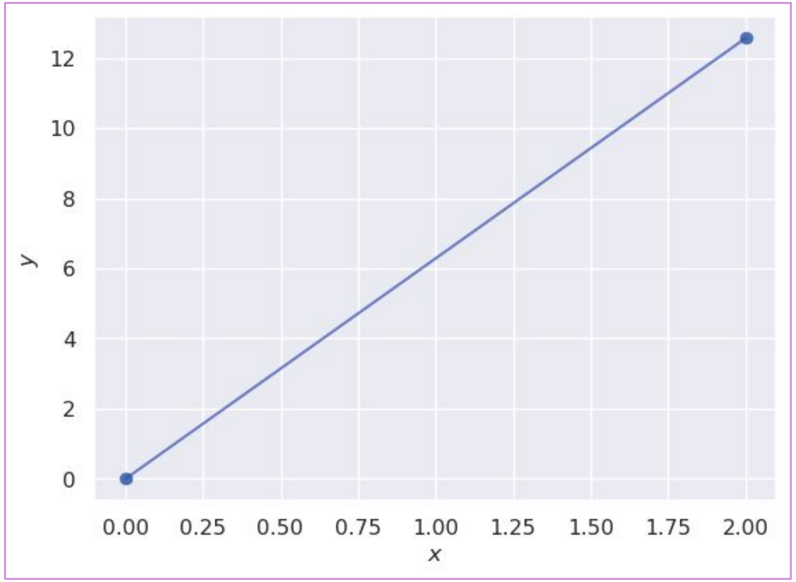
More on Calculus of Variations

- ➢ The arclength problem
- ➢ Brachistochrone problem
- ➢ Fermat’s principle
- ➢ Shape of a hanging rope

Statistical Methods

Find equation of the line which passes through the points: $(0, 0)$ and $(2, 4\pi)$ (slido only)



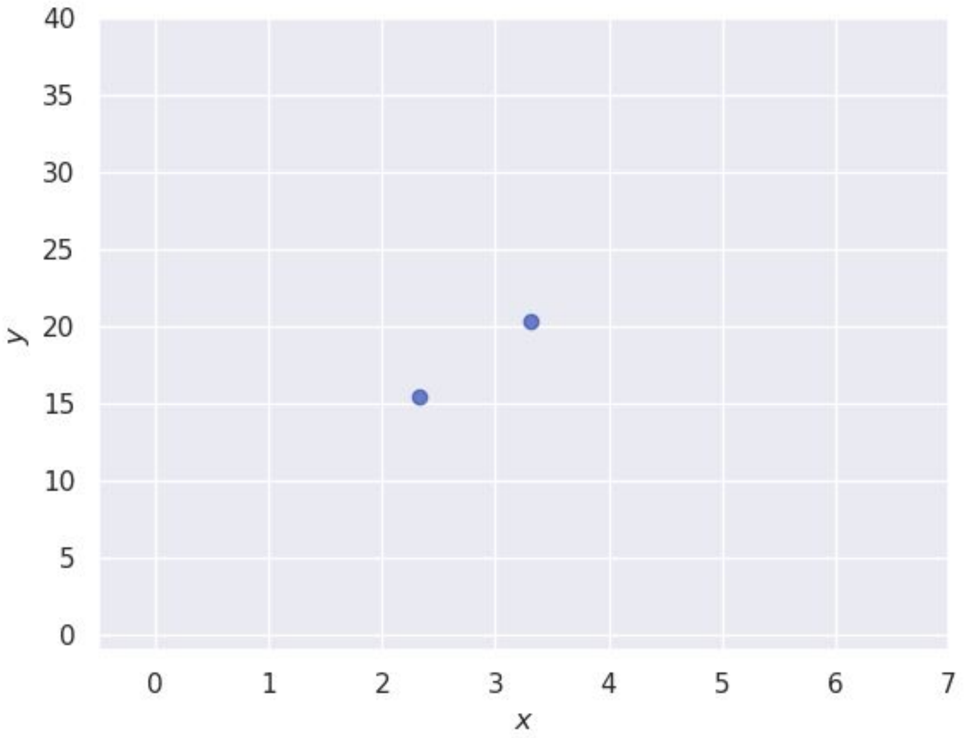


$$y = 2\pi x$$

Question (slido only)

As a mathematician, in one sentence, describe π

A Line Through n Points?



- Given a pencil, a ruler, and a pair of compass
- Draw many circles and measure
  - the circumference (C)
  - the radius (x)
- What is C/x ?
- Before π was discovered, nobody knew C/x is constant
- However, given a circle, one could easily measure its radius and circumference

Abundance of Data

Yes, Using Tools from Linear Algebra

$$\left.\begin{aligned} y_1 &= \xi_0 + \xi_1 x_1 \\ y_2 &= \xi_0 + \xi_1 x_2 \\ &\vdots \qquad\qquad \vdots \\ y_n &= \xi_0 + \xi_1 x_n \end{aligned}\right\} \Rightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix}$$

$\Rightarrow$ How to Solve $\boxed{A\xi = y}$?

Moore-Penrose Inverse (Newton Method)

- The Moore-Penrose Pseudo Inverse $A^+ = (A^T A)^{-1} A^T$ satisfies $\xi = A^+ y$ as a minimizer of the optimization problem

$$\underset{\xi \in \mathbb{R}^2}{\text{minimize}} \; \|A\xi - y\|^2$$

- A dual formulation of the minimization problem is

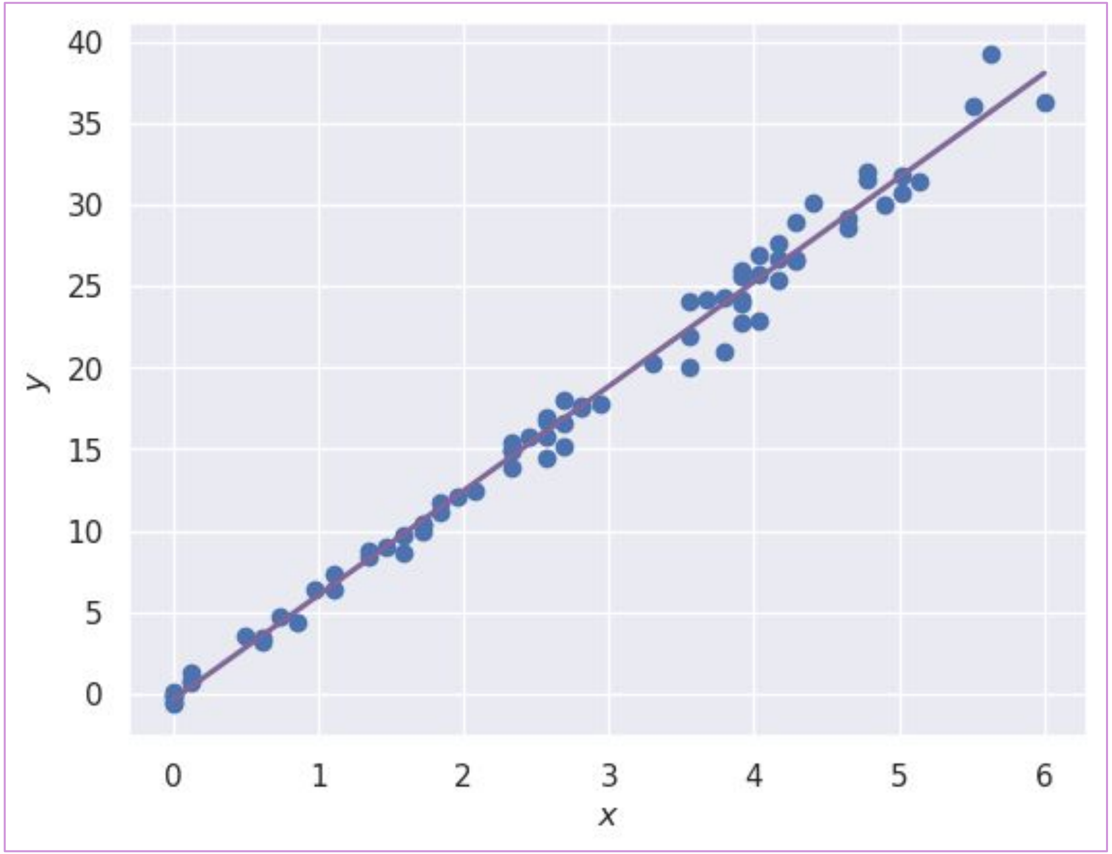
$$\underset{\xi \in \mathbb{R}^2}{\text{maximize}} \; p(y|\xi, x)$$

the maximum likelihood estimate, where

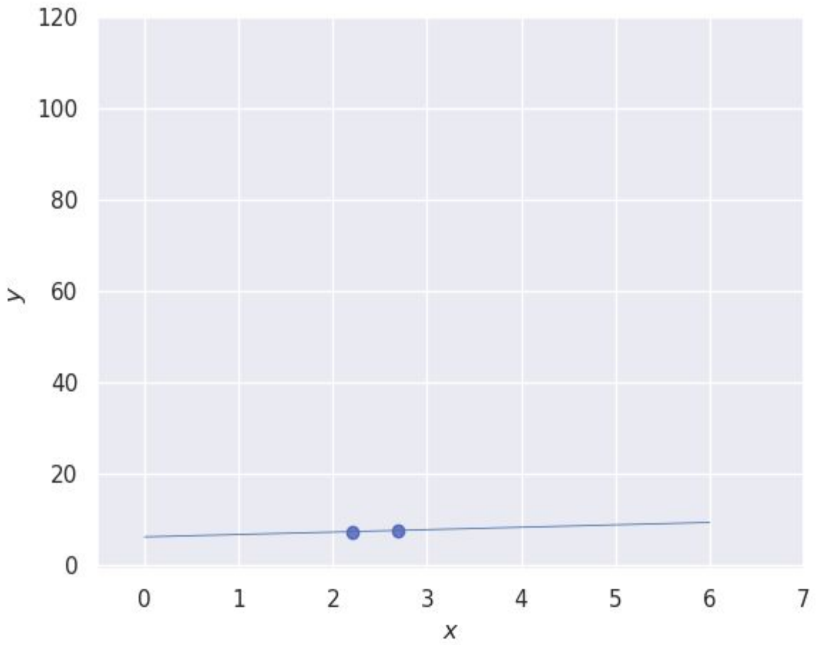
$$y = \xi^T x + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

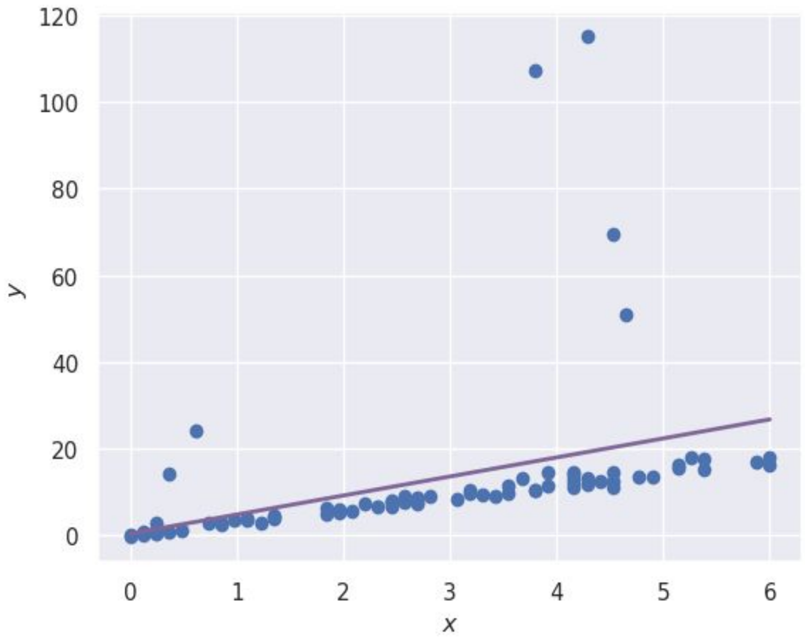
and x has 1 in its first dimension.

A Line Through n Points



Moore-Penrose Inverse Sensitive to Outliers





Methods of Solving Least Squares Problems

Linear Least Squares:

- Moore-Penrose Inverse
- Newton Method

Nonlinear Least Squares:

- Gradient Descent
- Gauss-Newton Method
- Levenberg-Marquardt method
- Stochastic Gradient Descent

Remarks

- Errors encountered in π estimation are mostly parallax error
- Parallax errors can be minimized by statistical averages, but pose uncertainties in measurements
- In heterogeneous media such as composites, geological media, gels, foams, and cell aggregates, these uncertainties could take any distribution, and in fact, could be undetermined useful material properties
- An accurate description of a measured value would as well characterize uncertainties in the obtained value

Differential Equations

- The Euler-Lagrange equation is a differential equation
- Rates are ubiquitous in day to day life
  - speed
  - acceleration
  - reaction rate
  - power
  - inflation rate
  - tax rate
  - unemployment rate
  - birth rate
  - interest rate
  - marginal
- More rates from Newton's laws and conservation laws in the natural and physical sciences

Functions From Differential Equations

➢ Consider the simple elliptic equation

$$\left.\begin{array}{c} -\dfrac{d}{dx}\left(\xi(x)\dfrac{dy}{dx}\right) = f(x), \quad x \in \Omega = [a,b] \\ y(a) = y_a,\ y(b) = y_b \end{array}\right\}$$

where ξ could be

- Young's modulus of a material
- Absolute permeability of rocks

Data Driven Modeling

- For any of these problems, ξ is never known but $\{(x_i, y_i)\}_{i=1}^n$ are easily, and in most cases, cheaply obtained
- Finding $y(x)$ from data is called data driven modeling
- In certain community, ξ is discovered through inverse problems
- In general, $\xi = \xi(x, y)$ may be heterogeneous

Regression

➢ Close your eyes to the physical law and fit

$$y(x) = \sum_{i=0}^{n} \xi_i \, \phi_i(x)$$

where ϕ_i ' s are elements of any suitable basis known to the researcher such as $\{1, x, x^2, \cdots, x^n\}$

➢ Suffers

- inductive bias
- futile adventure if solution lives outside the span of ϕ_i 's
- prior knowledge of physical laws are not exploited

➢ If solution lives in a subspace of the span of ϕ_i 's, techniques such as PCA are used to handle collinearity and dimensionality reduction

Weak Form (FEM)

➢ Let $v \in H^1(\Omega)$. Multiplying the differential form by v and integrating by parts gives

$$\int_\Omega \xi(x) y' \, v' dx = \int_\Omega f v dx \qquad \text{for all } v \in H^1(\Omega)$$

$$H^m(\Omega) := \{u \in L^2(\Omega) : \partial^i u \in L^2(\Omega) \quad \text{for all} \quad i \in [m]\}$$

Finite Element Approximation - Revisit

➢ Let V^h be a finite dimensional subspace of $H^1(\Omega)$ in which we seek an approximate solution of the form

$$y(x) \approx y_k(x) = \sum_{i=0}^{k} c_i \phi_i(x)$$

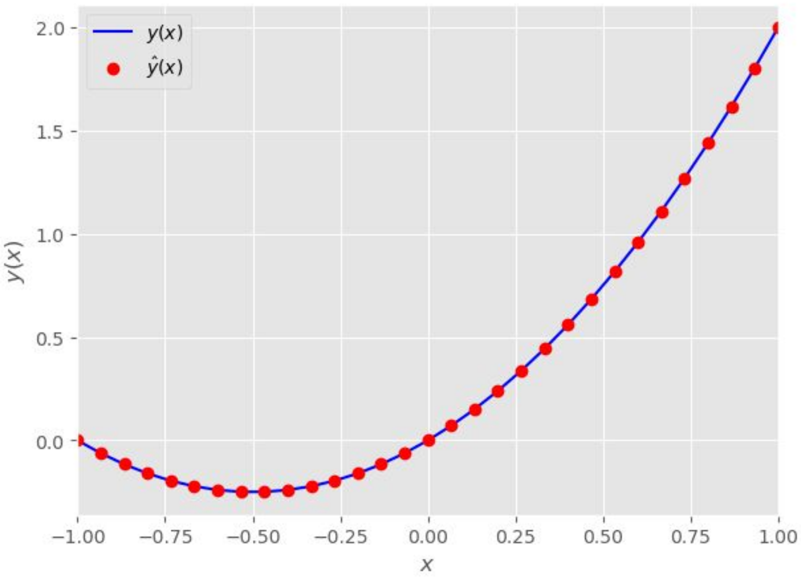
➢ Within the Galerkin framework, we assume $v \approx \phi_j \in V^h h$. So

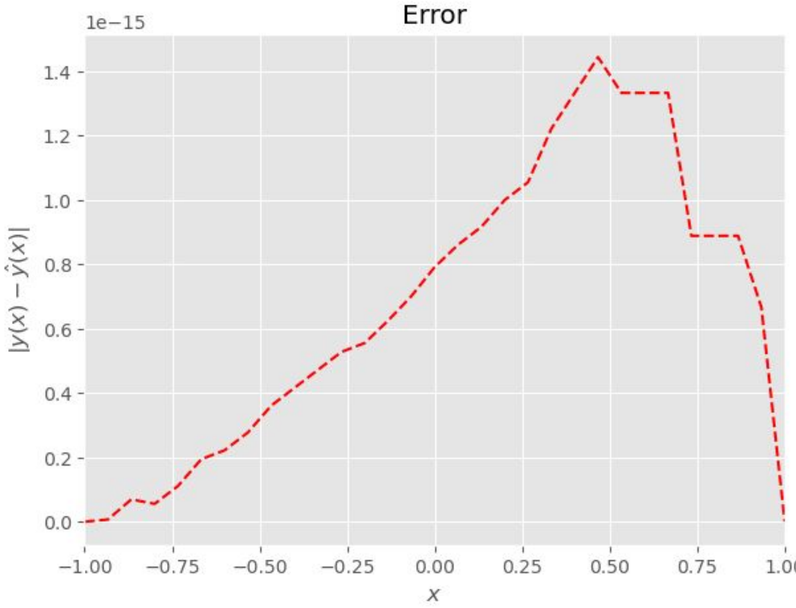
$$\sum_{i=0}^{k} c_i \int_\Omega \xi(x) \phi_i' \phi_j' dx = \int_\Omega f \phi_j dx \qquad \forall\ j \in [k]$$

simplifying to $Ac = b$ where $a_{ij} = \int_\Omega \xi(x) \phi_i' \phi_j' dx$ and $b_j = \int_\Omega f \phi_j dx$

Numerical Experiments

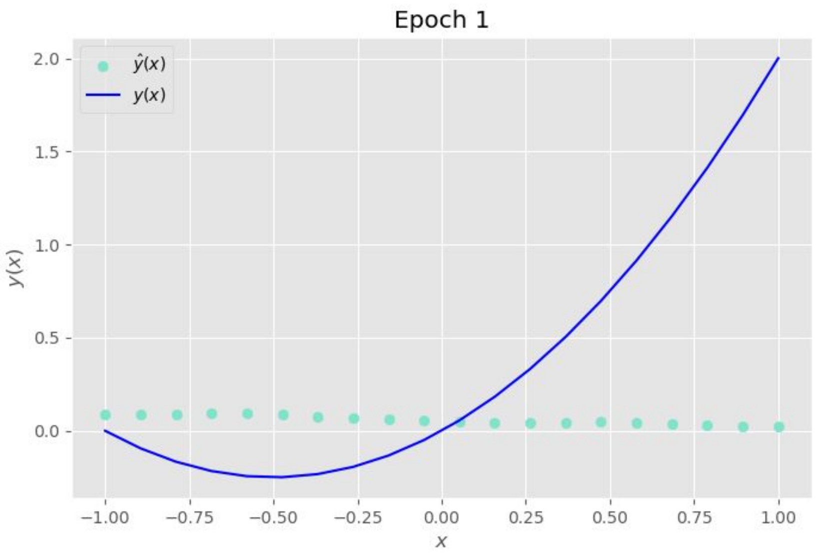
➢ With $\Omega = (-1,1)$ $\xi(x) = 1 + 0.5x$, $y(-1) = 0$, $y(1) = 2$, the load $f(x) = -(2.5 + 2x)$ and 30 elements

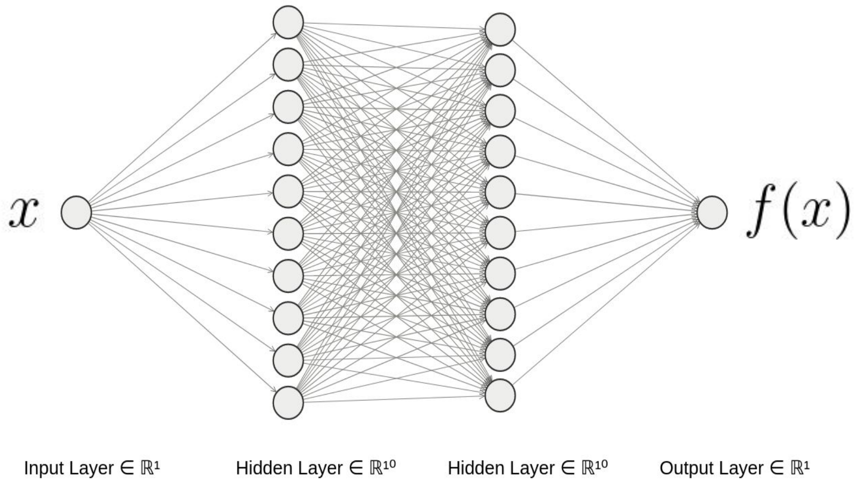




Neural Networks - Experiments

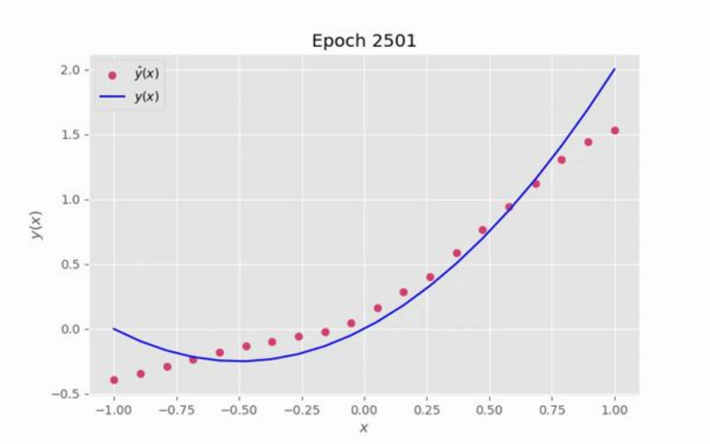
- ➢ We train a network of 3 fully connected layers, at 1000 sampled points, ReLU activation at the first two layers, MSE loss
- ➢ Adam optimizer, learning rate of 0.0001

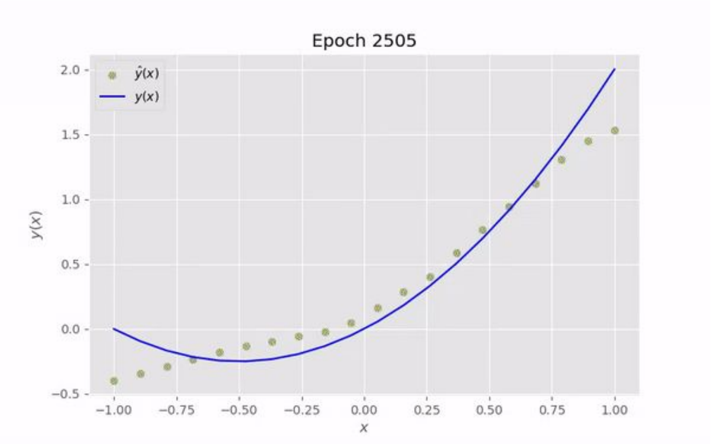


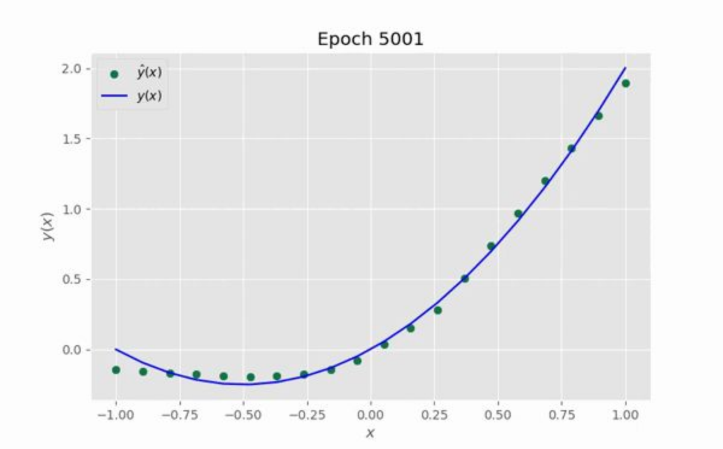



```
Net(
  (fc1): Linear(in_features=1, out_features=10, bias=True)
  (fc2): Linear(in_features=10, out_features=10, bias=True)
  (fc3): Linear(in_features=10, out_features=1, bias=True)
)
```

Convergence Requires High Epoch







Summary

- We presented a brief trajectory of functions in mathematics, from Euler-Lagrange to the state of the art machine learning models
- Gave insight on where the “least of the leasts” are applied in day to day life
- Showed how functions are discovered from data via statistics and differential equations
- Made connections between statistics, differential equations and calculus of variation
- Whether you are interested in pure or applied mathematics, you are stuck with functions
- The next time you think of pressing a button to get you a cup of coffee, I challenge you to think about the function behind the scene, no functions, no automation

Thanks for your attention