Function approximation

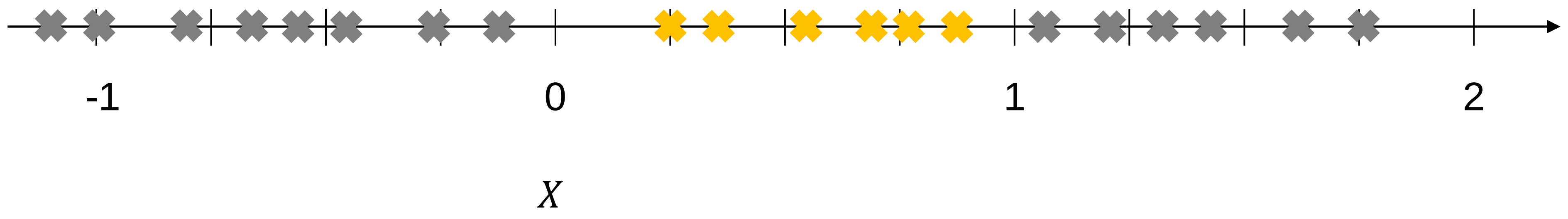
Lecture 13 of “Mathematics and AI”

Outline

1. A linear classifier for $y(x) = I[x^2 - x > 0]$
2. Feature mapping
3. Function approximations
   Taylor series, Fourier series, infinite-dimensional vector spaces
4. Koopman operator theory

A linear classifier for a quadratic function

Observations of $y(x) = I[x^2 - x > 0]$

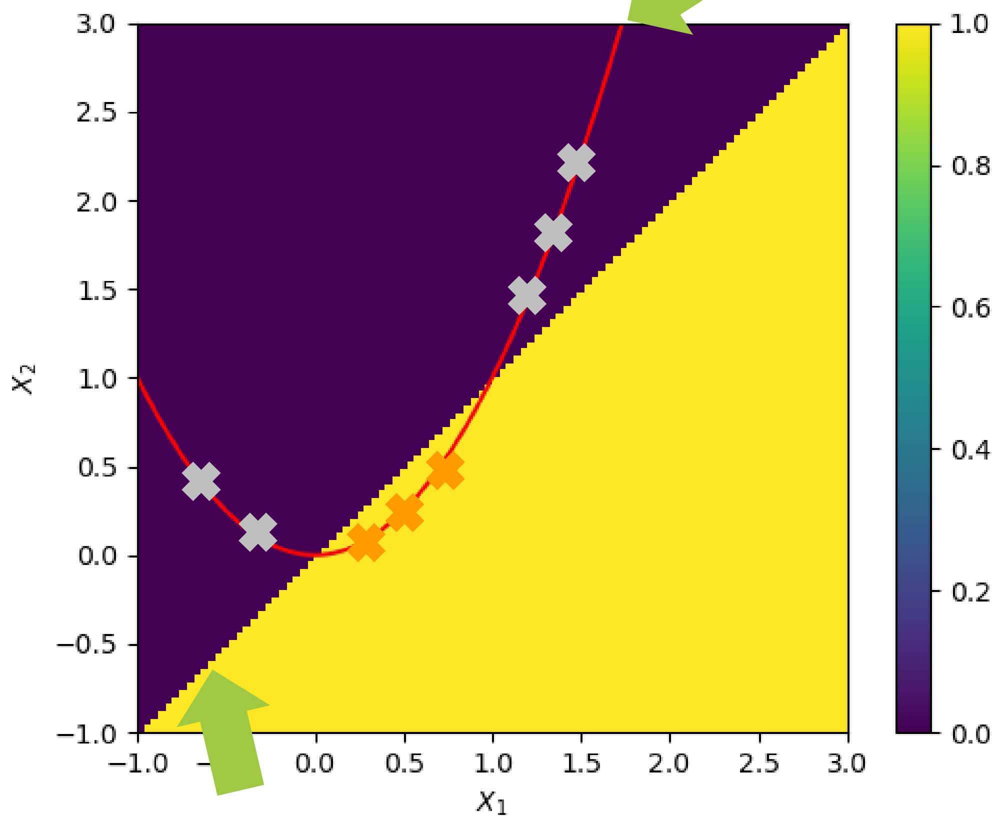


Manifold learning,
algebraic topology,
algebraic geometry

Observations of $y(x) = I[x_1^2 - x_1 > 0]$

- Introduce new variable x_2
- Define $x_2 := x_1^2$
  - $y(x) = I[x_2 - x_1 > 0]$
  - **Feature mapping/ Feature augmentation**

Manifold $x_2 = x_1^2$



Linear decision boundary $x_2 = x_1$

Feature mapping

DARTMOUTH

Feature mapping

- **How many** powers of x should we include?

- Should we use **other functions** of x?

- **What patterns** can be fitted in the augmented feature space?

Function approximations

Taylor's theorem and Taylor series

DEFINITION

If f has n derivatives at $x = a$, then the nth Taylor polynomial for f at a is

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

The nth Taylor polynomial for f at 0 is known as the nth Maclaurin polynomial for f.

- $p_k(x)$ is a ***linear combination*** of the ***functions*** $(x-a)^j$
- Each function $(x-a)^j$ in $p_k(x)$ has a **coefficient** $\beta := \frac{f^{(j)}(a)}{j!}$

THEOREM 6.7

Taylor's Theorem with Remainder

Let f be a function that can be differentiated $n+1$ times on an interval *I* containing the real number *a*. Let p_n be the *n*th Taylor polynomial of f at *a* and let

$$R_n(x) = f(x) - p_n(x)$$

be the *n*th remainder. Then for each *x* in the interval *I*, there exists a real number *c* between *a* and *x* such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

If there exists a real number *M* such that $\left|f^{(n+1)}(x)\right| \leq M$ for all $x \in I$, then

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}$$

for all *x* in *I*.

Function approximation with Taylor polynomials

THEOREM 6.8

Convergence of Taylor Series

Suppose that f has derivatives of all orders on an interval *I* containing *a*. Then the Taylor series

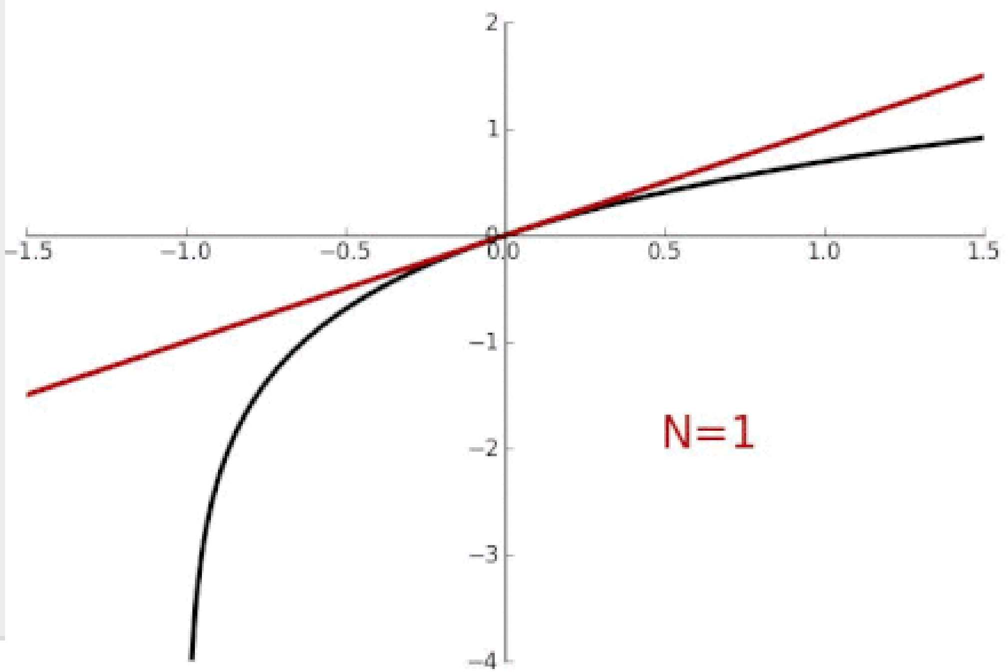
$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

converges to $f(x)$ for all *x* in *I* if and only if

$$\lim_{n\to\infty} R_n(x) = 0$$

for all *x* in *I*.

Example: $\ln(x+1)$



Taylor polynomials as models of finite complexity

- Consider (noise-free measurements of) a pattern
  - $y(x) = f(x)$ (regression) or
  - $y(x) = I[f(x) > 0]$ (classification)
- If $f(x)$ is an order-k polynomial
  - ➢ $p_k(x)$ is a perfect model
- If $f(x)$ can be approx. by an order-k polynomial with a small error
  - ➢ $p_k(x)$ is a good model

Taylor series as models of infinite complexity

Function	Maclaurin Series	Interval of Convergence
$f(x) = \frac{1}{1-x}$	$\sum_{n=0}^{\infty} x^n$	$-1 < x < 1$
$f(x) = e^x$	$\sum_{n=0}^{\infty} \frac{x^n}{n!}$	$-\infty < x < \infty$
$f(x) = \sin x$	$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$	$-\infty < x < \infty$
$f(x) = \cos x$	$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$	$-\infty < x < \infty$
$f(x) = \ln(1+x)$	$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}$	$-1 < x \leq 1$
$f(x) = \tan^{-1} x$	$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$	$-1 \leq x \leq 1$
$f(x) = (1+x)^r$	$\sum_{n=0}^{\infty} \binom{r}{n} x^n$	$-1 < x < 1$

Source: OpenStax Calculus 2

Fourier series

- Taylor series

$$f(x) = \sum_{j=1}^{\infty} \beta_j x^j$$

has "∞" coefficients

$$\beta_j = \frac{f^{(j)}(a)}{j!}$$

- For polynomial f, only a finite number of $\beta_j \neq 0$

Fourier series

- Fourier series

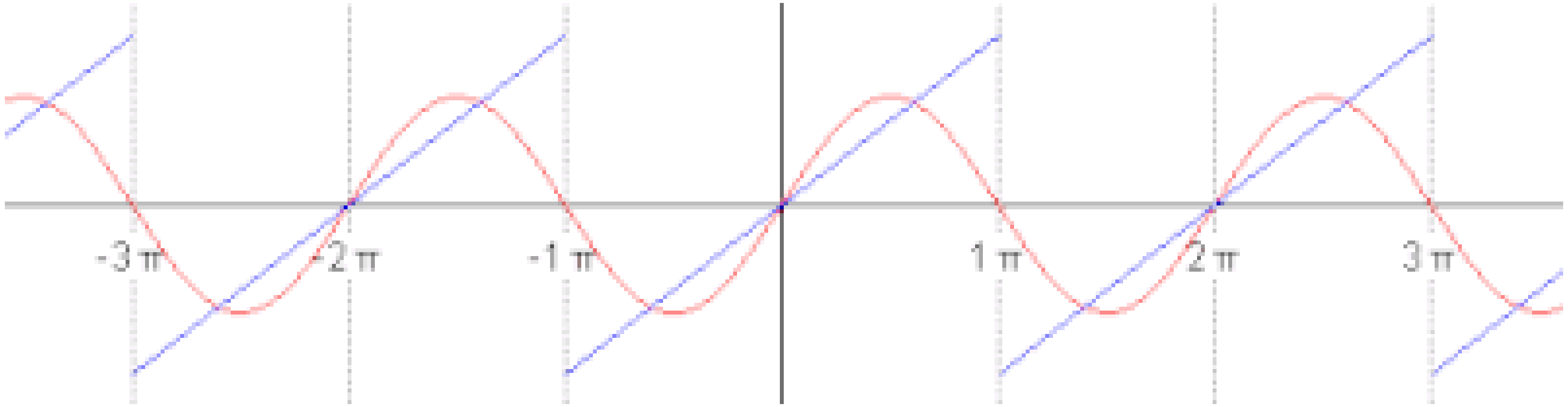
$$f(x) = \frac{1}{2}\alpha_0 + \sum_{j=1}^{\infty} \alpha_j \cos(jx) + \sum_{j=1}^{\infty} \beta_j \sin(jx)$$

has "$2\infty + 1$" coefficients

$$\alpha_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx\,, \qquad \alpha_j = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\cos(jx)\,dx\,, \qquad \beta_j = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\sin(jx)\,dx$$

- relevant for applications in signal processing of periodic signals

Fourier series



Vector spaces of functions

DEFINITION

A vector is a quantity that has both magnitude and direction.

THEOREM 2.1

Properties of Vector Operations

Let $\mathbf{u}, \mathbf{v},$ and $\mathbf{w}$ be vectors in a plane. Let r and s be scalars.

i.	$\mathbf{u} + \mathbf{v}$	$=$	$\mathbf{v} + \mathbf{u}$	Commutative property
ii.	$(\mathbf{u} + \mathbf{v}) + \mathbf{w}$	$=$	$\mathbf{u} + (\mathbf{v} + \mathbf{w})$	Associative property
iii.	$\mathbf{u} + \mathbf{0}$	$=$	$\mathbf{u}$	Additive identity property
iv.	$\mathbf{u} + (-\mathbf{u})$	$=$	$\mathbf{0}$	Additive inverse property
v.	$r(s\mathbf{u})$	$=$	$(rs)\mathbf{u}$	Associativity of scalar multiplication
vi.	$(r+s)\mathbf{u}$	$=$	$r\mathbf{u} + s\mathbf{u}$	Distributive property
vii.	$r(\mathbf{u} + \mathbf{v})$	$=$	$r\mathbf{u} + r\mathbf{v}$	Distributive property
viii.	$1\mathbf{u}$	$=$	$\mathbf{u}, 0\mathbf{u} = \mathbf{0}$	Identity and zero properties

Definition: A **vector space** consists of a set V (elements of V are called **vectors**), a field $\mathbb{F}$ (elements of $\mathbb{F}$ are called **scalars**), and two operations

- An operation called *vector addition* that takes two vectors $v, w \in V$, and produces a third vector, written $v + w \in V$.
- An operation called *scalar multiplication* that takes a scalar $c \in \mathbb{F}$ and a vector $v \in V$, and produces a new vector, written $cv \in V$.

which satisfy the following conditions (called *axioms*).

1. Associativity of vector addition: $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$.
2. Existence of a zero vector: There is a vector in V, written 0 and called the **zero vector**, which has the property that $u + 0 = u$ for all $u \in V$
3. Existence of negatives: For every $u \in V$, there is a vector in V, written $-u$ and called the **negative of** u, which has the property that $u + (-u) = 0$.
4. Associativity of multiplication: $(ab)u = a(bu)$ for any $a, b \in \mathbb{F}$ and $u \in V$.
5. Distributivity: $(a + b)u = au + bu$ and $a(u + v) = au + av$ for all $a, b \in \mathbb{F}$ and $u, v \in V$.
6. Unitarity: $1u = u$ for all $u \in V$.

Source: OpenStax Calculus 2, https://www.math.toronto.edu/gscott/WhatVS.pdf

Examples of function spaces and bases

- C^k and C^∞
  - Definition: the space of k -times differentiable functions
  - Possible basis: polynomials (Taylor series)
- L^p spaces, Sobolov spaces
  - Intuition: spaces of nice* functions that taper off quickly* as $x \to \pm\infty$
  - Possible basis: Sines and cosines (Fourier series)
    or radial basis functions (e.g., $\exp(-rc_j)$)

*lots of technical precision omitted here for the sake of intuitive understanding

Koopman operator theory

II. The Koopman Operator

Let $\mathcal{D} \subseteq \mathbb{R}^n$, let $f\colon \mathcal{D} \to \mathcal{D}$, and, for all $k \geq 0$, consider the discrete-time system

$$x(k+1) = f(x(k)), \quad x(0) = x_0, \tag{1}$$

where $x_0 \in \mathbb{R}^n$ is the initial condition. In addition, let $g\colon \mathcal{D} \to \mathbb{R}$ determine the output of (1) given by

$$y(k) = g(x(k)). \tag{2}$$

Note that, for all $k \geq 0$,

$$g(x(k)) = g(f^{(\circ k)}(x_0)) = (g \circ f^{(\circ k)})(x_0), \tag{3}$$

where $f^{(\circ 0)}$ is the identity map on $\mathcal{D}$, $f^{(\circ 1)} \triangleq f$, $f^{(\circ 2)} \triangleq f \circ f$, and $f^{(\circ k)}$ denotes the composition of f with itself k times.

Next, motivated by (3), let $\mathcal{G}$ be a set of real-valued functions on $\mathcal{D}$ which satisfy the following property:

$$\text{If } g \in \mathcal{G}, \text{ then } g \circ f \in \mathcal{G}. \tag{4}$$

If (4) holds, then $\mathcal{G}$ is *compositionally complete*. Note that, if $g \in \mathcal{G}$, then, for all $r \geq 0$, $g \circ f^{(\circ r)} \in \mathcal{G}$. If $\mathcal{G}$ is compositionally complete and $g \in \mathcal{G}$, then g is an *observable*.

Next, define the *Koopman operator* $\mathcal{K}\colon \mathcal{G} \to \mathcal{G}$ by

$$\mathcal{K}(g) \triangleq g \circ f. \tag{5}$$

Note that, since $\mathcal{G}$ is compositionally complete, it follows that, if $g \in \mathcal{G}$, then $g \circ f \in \mathcal{G}$ and thus $\mathcal{K}(g) \in \mathcal{G}$.

We now assume that $\mathcal{G}$ is a compositionally complete vector space over $\mathbb{R}$. Then, for all $g_1, g_2 \in \mathcal{G}$ and all $a_1, a_2 \in \mathbb{R}$,

$$\begin{aligned} \mathcal{K}(a_1 g_1 + a_2 g_2) &= (a_1 g_1 + a_2 g_2) \circ f \\ &= a_1 g_1 \circ f + a_2 g_2 \circ f \\ &= a_1 \mathcal{K}(g_1) + a_2 \mathcal{K}(g_2). \end{aligned} \tag{6}$$

Therefore, $\mathcal{K}$ is a linear operator on $\mathcal{G}$.

In terms of powers of $\mathcal{K}$, note that, for all $g \in \mathcal{G}$, $\mathcal{K}^0(g) = g \circ f^{\circ 0} = g$, thus $\mathcal{K}^0$ is the identity map on $\mathcal{G}$. Furthermore, for all $g \in \mathcal{G}$, $\mathcal{K}^2(g) = \mathcal{K}(\mathcal{K}(g)) = \mathcal{K}(g \circ f) = (g \circ f) \circ f = g \circ (f \circ f) = g \circ f^{\circ 2}$, and thus, for all $k \geq 0$, $\mathcal{K}^k(g) = g \circ f^{\circ k}$. Hence, for all $k \geq 1$, the output (2) is given by

$$y(k) = \mathcal{K}^k(g)(x_0) = \mathcal{K}^{k-1}(g)(x_1). \tag{7}$$