Computational Statistics & Machine Learning

Lecture 9

Sobolev spaces and Function approximation

Mark Girolami
mag92@cam.ac.uk

Department of Engineering
University of Cambridge

November 3, 2021

Overview

- Subspace of Banach space - characterising smoothness
- Weak derivatives
- Function approximation by Polynomial expansion
- Understanding rate of convergence

Sergei Lvovich Sobolev

- Sergei Lvovich Sobolev 1908 - 1989

What is a Sobolev Space?

- The definition of the norm is a combination of L^p norms of function and its derivatives up to a given order
- The derivative used is a *weak derivative*
- Sobolev spaces important in the solution of partial differential equations
- Important in Machine Learning for function approximation and assessing generalisation performance
- We shall work through an illustrative example

What is a Sobolev Space?

- Choose a space of smooth continuous functions of degree k (possessing k derivatives)
- The functions - in this simple case - are defined on a domain which is an open set of the real line $\Omega \subset \mathbb{R}$
- Denote this space as $C^k(\Omega)$ now define the norm as

$$\begin{aligned}\|f\|_{W_p^k}^p = \int_\Omega |f(x)|^p dx &+ \int_\Omega |Df(x)|^p dx \\ &+ \int_\Omega |D^2 f(x)|^p dx \\ &\vdots \\ &+ \int_\Omega |D^k f(x)|^p dx\end{aligned}$$

- So

$$\|f\|_{W_p^k} = \left(\sum_{s=0}^{k} \int_\Omega |D^s f(x)|^p dx\right)^{\frac{1}{p}}$$

What is a Sobolev Space?

- This definition does not provide for a complete space as seen in previous lecture
- The limit in $C^k(\Omega)$ may extend beyond the scope of the norm $\|f\|_{W_p^k}$
- Enlarge to a Lebesgue space such that the following definition holds
- Consider a function $f \in L^p(\Omega)$ defined on $\Omega \subset \mathbb{R}$
- Let f have weak derivatives up to k'th degree which also belong to $L^p(\Omega)$
- A Sobolev space $W_p^k(\Omega)$ is defined by the set of functions

$$W_p^k(\Omega) = \{f \in L^p(\Omega); D^s f \in L^p(\Omega), 0 \leq s \leq k\}$$

with norm

$$\|f\|_{W_p^k} = \left(\sum_{s=0}^{k} \int_\Omega |D^s f(x)|^p dx\right)^{\frac{1}{p}}$$

What is a Sobolev Space?

- For the case where $p = 2$ then $W_2^k(\Omega)$ is a Hilbert space with inner-product

$$\langle f, g \rangle_{W_2^k(\Omega)} = \sum_{s=0}^{k} \langle D^s f, D^s g \rangle$$

- One can observe that both the norm and inner-product now encode aspects of the *smoothness* of the function classes we shall see why this is useful in Machine Learning shortly
- The weak derivative is hugely important in the solution of PDE's and forms the basis of the finite element method
- We will introduce the weak derivative in a way to gain intuition rather than clutter with details

Weak Derivative

- The concept of the weak derivative is related to the generalisation of the definition of functions to *distributions*
- Consider the function f defined on the interval $[0, 2]$ such that

$$f(x) = \begin{cases} x & \text{for } \; x \in [0, 1) \\ 1 & \text{for } \; x \in [1, 2] \end{cases}$$

- The function at the troublesome point at $x = 1$ does not have a derivative
- If this could be ignored the derivative would be defined as

$$f'(x) = \begin{cases} 1 & \text{for } \; x \in [0, 1) \\ 0 & \text{for } \; x \in [1, 2] \end{cases}$$

Weak Derivative

- In simple terms consider two functions defined on the unit line $[0,1]$ where the function g take zero values at the end points $g(0) = g(1) = 0$
- As the boundary terms are zero then schoolboy integration by parts tells us that

$$\int_0^1 f(x)g'(x)dx = -\int_0^1 f'(x)g(x)dx$$

- Exchanging differentiability of f for that of g
- Let $f : [0,1] \to \mathbb{R}$ be a real valued function, the function $h : [0,1] \to \mathbb{R}$ is a *weak derivative* of f if for every differentiable function $g : [0,1] \to \mathbb{R}$ with $g(0) = g(1) = 0$ it holds that

$$\int_0^1 f(x)g'(x)dx = -\int_0^1 h(x)g(x)dx$$

- Generalising to multiple dimensions over more general domains and higher derivatives follows straightforwardly albeit with an increase in tedious notation

Weak Derivative

- Consider the function f defined on the interval $[0,2]$ such that

$$f(x) = \begin{cases} x & \text{for } \; x \in [0,1) \\ 1 & \text{for } \; x \in [1,2] \end{cases}$$

- For any differentiable test function $\phi : [0,2] \to \mathbb{R}$, with $\phi(0) = \phi(2) = 0$ then

$$\begin{aligned} -\int_0^1 x\phi'(x)dx &= -x\phi(x)|_0^1 + \int_0^1 \phi(x)dx \\ &= -\phi(1) + \int_0^1 \phi(x)dx \end{aligned}$$

Weak Derivative

- Furthermore

$$-\int_1^2 \phi'(x)dx = -\phi(2) + \phi(1) = \phi(1)$$

- Finally

$$-\int_0^2 f(x)\phi'(x)dx = \int_0^1 \phi(x)dx = \int_0^2 g(x)\phi(x)dx$$

- Where g is

$$g(x) = \begin{cases} 1 & \text{for } x \in [0,1) \\ 0 & \text{for } x \in [1,2] \end{cases}$$

- The weak derivative we intuitively wanted, a non-continuous function defined on sets up to measure zero

Function approximation by Polynomial expansion

- One of the main applications in Machine Learning is learning functions from data
- The key question is how fast learning takes place, in other words at what rate does the approximation error reduce and go to zero
- We will use a very simple example to obtain insight and illustrate the analysis
- The same type of analysis can be used for Deep Learning and more complex models and functions

Function approximation by Polynomial expansion

- Consider functions from the set

$$C_2[-\pi,\pi] = C[-\pi,\pi] \cap L^2[-\pi,\pi]$$

- What type of functions does this set contain?
- Fourier expansions can be used to represent functions from this set

$$f(x) = \sum_{k=0}^{\infty} c_k \exp(ikx),\ c_k \propto \int_{-\pi}^{\pi} f(x)\exp(-ikx)dx$$

- From previous lecture we know that the L^2 norm is

$$\|f\|_{L^2}^2 = \sum_{k=0}^{\infty} c_k^2$$

Function approximation by Polynomial expansion

- The space of approximating functions will be the Sobolev space

$$W_2^s = \left\{ f \in C_2[-\pi, \pi]; \|D^s f\|_{L^2}^2 < +\infty \right\}$$

- It (should be) easy to see that the norm in this space follows as

$$\|f\|_{W_2^s}^2 = \|D^s f\|_{L^2}^2 = \sum_{k=1}^{\infty} k^{2s} c_k^2$$

- For this to be bounded then the Fourier coefficients need to decay at a rate which increases with s what does this mean ?

Function approximation by Polynomial expansion

- Let us choose an increasing space H_n as the set of trigonometric polynomials of degree n

$$p(x) = \sum_{k=1}^{n} a_k \exp(ikx)$$

- So for a function of the form we are considering

$$f(x) = \sum_{k=0}^{\infty} c_k \exp(ikx)$$

it is clear the optimal approximation is

$$f_n(x) = \sum_{k=1}^{n} c_k \exp(ikx)$$

Function approximation by Polynomial expansion

- For an $f \in W_2^s$ the approximation error

$$\epsilon_n[f] = \|f - f_n\|_{L^2}^2$$

  needs to be studied
- For an n-degree polynomial there will be n parameters to be estimated in the approximation
- It is clear that as $n \to \infty$ then $\epsilon_n \to 0$ very good
- The real question is *how fast* does this happen , how quickly do we learn

Function approximation by Polynomial expansion

- Use a number of bounds

$$
\begin{aligned}
\epsilon_n[f] &= \|f - f_n\|_{L^2}^2 = \sum_{k=n+1}^{\infty} c_k^2 = \sum_{k=n+1}^{\infty} c_k^2 \times k^{2s}\frac{1}{k^{2s}} \\
&< \frac{1}{n^{2s}} \sum_{k=n+1}^{\infty} k^{2s} c_k^2 < \frac{1}{n^{2s}} \sum_{k=1}^{\infty} k^{2s} c_k^2 \\
&= \frac{1}{n^{2s}} \|f\|_{W_2^s}^2
\end{aligned}
$$

- We have that

$$
\epsilon_n[f] < \frac{1}{n^{2s}} \|f\|_{W_2^s}^2
$$

- As s increases the impact of increasing n is greater , the error decreases faster as the model complexity n is increased
- The rate of convergence is faster as the smoothness of the function being approximated increases

Function approximation by Polynomial expansion

- What happens in more than one dimension?
- If the function is defined in a d-dimensional domain such that for $d = 2$ then

$$p(x) = \sum_{k,m=1}^{l} a_{k,m} \exp(ikx + imy)$$

- Show that

$$\epsilon_n[f] < \frac{1}{2n^{\frac{2s}{d}}} \|f\|^2_{W_2^s}$$

- What can we infer from this inequality ?
- Clearly as function smoothness increases convergence rates increase
- As dimension increases convergence rate decreases - the curse of dimensionality

Function approximation by Polynomial expansion

- This type of analysis is used to study Deep Nets, Gaussian Processes, and other function approximations
- The basic concepts are the same it just gets a bit more technical
- Understanding the function spaces being considered and their properties is essential to developing effective Machine Learning methods
- We have looked at a number of characteristics of function spaces that prove useful
- We now need to think about measures (Lebesgue, Probability) on function space.