Section 3: Polynomial interpolation

February 28, 2019

1 Polynomial interpolation: the basics

Let us look at the following problem. Suppose we are given the data points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n). Can we find a polynomial which passes through exactly all the points? Such an polynomial is called an interpolating polynomial. Let

$$P_{n-1}(x) = a_1 + a_2 x + a_3 x^2 + \ldots + a_n x^{n-1}$$

denote a polynomial of degree $n-1$. The degree is equal to $n-1$ since the highest power is equal to $n-1$. It requires n coefficients to uniquely describe a polynomial of degree $n-1$. This is because the terms x^k (also called monomials) for $k = 1, 2, \ldots, n-1$ are linearly independant from each other. After all, there exist no coefficients c_1, c_1,..., c_{k-1} for which $x^k = c_1 + c_2 x + c_3 x^2 + \ldots + c_k x^{k-1}$.

Given this observation, one may wonder that a polynomial of at least degree $n-1$ is needed to pass through all n data points. This answer is correct under some additional conditions. The best way to show this, is to go on and attempt at solving for the interpolating polynomial. We have the following equations to look at:

$$\begin{aligned}
a_1 + a_2x_1 + a_3x_1^2 + \ldots + a_nx_1^{n-1} &= y_1\\
a_1 + a_2x_2 + a_3x_2^2 + \ldots + a_nx_2^{n-1} &= y_2\\
&\vdots\\
a_1 + a_2x_n + a_3x_n^2 + \ldots + a_nx_n^{n-1} &= y_n
\end{aligned}$$

In matrix notation, this can be expressed as:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1}\\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} a_1\\ a_2\\ \vdots\\ a_n \end{bmatrix} = \begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_n \end{bmatrix}$$

The matrix above is called the vandermonde matrix, the vandermonde matrix associated with the monomial basis to precise. This matrix is square, but we do not know if it invertible. If it is, we have found the unique interpolating polynomial of degree $n-1$. Uniqueness follows from the fact that the matrix inverse is unique, and from our earlier discussion concerning linear independence of the monomial terms. The vandermonde matrix is invertible under very mild conditions.

Theorem 1. *The vandermonde matrix:*

$$V_{n-1}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1}\\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

is invertible if, and only if:

$$x_i \neq x_j, \quad \textit{whenever } i \neq j.$$

Proof. A matrix is invertible if, and only if, its determinant is non-zero. The determinant of the $V_{n-1}(x_1, x_2, \ldots, x_n)$ turns out to be equal to:

$$\det V_{n-1}(x_1, x_2, \ldots, x_n) = \prod_{1\leq i<j\leq n} (x_j - x_i) = \prod_{j=1}^{n} \left(\prod_{i=1}^{j-1} (x_j - x_i) \right)$$

From here, the condition $x_i \neq x_j, \quad$ whenever $i \neq j$ follow trivially. The hard part is to show that the above expression for the determinant is correct. $\square$

Lagrange interpolation formula. Inversion of the vandermonde matrix requires work, and is not the best way to find the unique interpolating polynomial. Lucky for us, a polynomial does not always have to be expressed in terms of the monomial terms x^k. Let's look at the following construction:

$$P(x) = y_1 \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + y_2 \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + y_3 \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}.$$

Notice when $x = x_1$, we see that the second and third term are equal to zero:

$$\left.\frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)}\right|_{x=x_1} = 0, \qquad \left.\frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}\right|_{x=x_1} = 0$$

whereas the first term equals:

$$\left.\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}\right|_{x=x_1} = 1$$

This yields $P(x_1) = y_1$. Likewise, $P(x_2) = y_2$ and $P(x_3) = y_3$.

This clever construction can be generalized and is called the Lagrange interpolation formula. Call:

$$L_{n-1,k}(x) = \prod_{j=1, j\neq k}^{n} \frac{x - x_j}{x_k - x_j}$$

and notice that:

$$L_{n-1,k}(x_l) = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

Expressing our interpolating polynomial as:

$$P_{n-1}(x) = \sum_{k=1}^{n} a_k L_{n-1,k}(x)$$

The vandermonde matrix in terms of the Lagrange polynomials is equal to:

$$\begin{bmatrix} L_{n-1,1}(x_1) & L_{n-1,2}(x_1) & L_{n-1,3}(x_1) & \cdots & L_{n-1,n}(x_1) \\ L_{n-1,1}(x_2) & L_{n-1,2}(x_2) & L_{n-1,3}(x_2) & \cdots & L_{n-1,n}(x_2) \\ \vdots & \vdots & \vdots & & \vdots \\ L_{n-1,1}(x_n) & L_{n-1,2}(x_n) & L_{n-1,3}(x_n) & \cdots & L_{n-1,n}(x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

We see that $a_k = y_k$, and voila:

$$P_{n-1}(x) = \sum_{k=1}^{n} y_k L_{n-1,k}(x) \tag{1}$$

is our interpolating polynomial.

Newton divided difference. There are other ways of describing the interpolating polynomial apart from (1). Newton's divided difference formula is one of them. We won't be covering them during this course.

2 Approximating functions through polynomial interpolation

Let us bring a new dimension to our interpolation problems. Suppose that the values y_1, y_2, ...,y_n are obtained from some continuous function $f(x)$. That is, $y_1 = f(x_1)$, $y_2 = f(x_2)$, ..., $y_n = f(x_n)$. We can ask some interesting questions from here. Most notably, is it generally true that as we sample more points from the interval $[a, b]$, the resulting interpolating polynomial becomes a better approximation of the original function. As you may imagine, such a result can be extremely useful. Polynomials are, for example, extremely nice objects for differentation and integration. If we can use polynomials to approximate other more complicated functions, we can use them as a proxy to differentiate and integrate those complicated functions approximatively.

Weierstrauss's result. One question we must ask beforehand is whether it is even possible to approximate arbitrary function through polynomial. We have the classic result, due to Weierstrauss.

Theorem 2. *Let $f(x)$ be a continuous function on $[a, b]$. And let $P_n(x)$ denote a polynomial of degree n, then there exists a sequence P_1, P_2, ... such that:*

$$\lim_{n\to\infty} \max_{x\in[a,b]} |f(x) - P_n(x)| = 0$$

Proof. There are several ways to prove this. One constructive proof is through Bernstein polynomials, which will be one of the extra homework problems during this course. □

The theorem above does not imply necessarily an interpolating polynomial, which is just one of way constructing a polynomial approximation to a function. The theory of Bernstein polynomials presents another way of approximating polynomial. Later on, we will discuss how one can find the best approximating polynomial in the 2-norm. For now, just keep in mind that the subject of approximation theory is a vast one, and we only scratched the surface.

Lagrange error formula. Coming back to our discussion on interpolating polynomials. Let us move on and derive an useful formula which can be used to describe the error between the interpolating polynomial and the function itself.

Theorem 3. *Let $f(x)$ be a n times differentiable function on $[a, b]$ and let $P_{n-1}(x)$ be the interpolating polynomial of degree $n - 1$ with interpolation points $a \le x_1 < x_2 < \ldots < x_n \le b$. For every $\xi \in [a, b]$, then there exists some $\eta \in [a, b]$ for which:*

$$f(\xi) = P_{n-1}(\xi) + \frac{f^{(n)}(\eta)}{n!} \prod_{k=1}^{n} (\xi - x_k)$$

Proof. Notice that we may re-write $f(x)$ as:

$$f(x) = P(x) + M(x)\prod_{k=1}^{n}(x - x_k), \quad M(x) = \frac{f(x) - P(x)}{\prod_{k=1}^{n}(x - x_k)}.$$

Based on that, introduce:

$$\hat{P}(x) = P_{n-1}(x) + M(\xi)\prod_{k=1}^{n}(x - x_k)$$

By construction, at $x = \xi$, we have $\hat{P}(\xi) = f(\xi)$. However, in general:

$$\begin{aligned} g(x) &= f(x) - \hat{P}(x) \\ &= f(x) - P_{n-1}(x) - M(\xi)\prod_{k=1}^{n}(x - x_k). \end{aligned}$$

Let us make some observations about the function $g(x)$:

(i) $g(x_1) = g(x_2) = \ldots = g(x_n) = 0$ and $g(\xi) = 0$.

(ii) $g^{(n)}(x) = f^{(n)}(x) - n!M(\xi)$

What we will do next is show that there exists some $\eta \in [a, b]$ for which $g^{(n)}(\eta) = 0$. From (ii), this establishes $M(\xi) = \frac{f^{(n)}(\eta)}{n!}$, subsequently proving the theorem. The following lemma, also called Rolle's theorem, will come in handy.

lemma 1. *Let $f(x)$ be a continuously differentiable function, i.e. the derivative $f'(x)$ exists and is a continuous function. Suppose that $f(a) = 0$ and $f(b) = 0$ with $b > a$. Then, there exists an $c \in [a, b]$ for which $f'(c) = 0$.*

Now looking at (i), we can apply the above lemma repetitively to obtain what we desire. The function $g(x)$ has at least $n+1$ roots. Using the lemma, $g'(x)$ has at least n roots. Then $g^{(2)}(x)$ has at least $n-1$ roots. As we keep going, $g^{(n)}(x)$ at least one root, call it η. $\square$

The nice thing about the above theorem is that we can use bounds on the derivatives of f to establish bounds on the interpolation error. Indeed,

$$|f(x) - P_{n-1}(x)| \leq \max_{\eta \in [a,b]} \left| \frac{f^{(x)}(\eta)}{n!} \right| \max_{x \in [a,b]} \prod_{k=1}^{n} |x - x_k|$$

In the above inequality, notice we have no control over the derivatives of the function. That is fixed. However, we do have freedom in choosing the interpolation points. The question now is, how should we select the interpolation so that we minimize:

$$\max_{x \in [a,b]} \prod_{k=1}^{n} |x - x_k| \,.$$

Chebyshev polynomials and Chebyshev nodes To answer this question, we must talk about Chebyshev polynomials. Chebhyshev polynomials (of the first kind) are defined by:

$$T_k(x) = \cos(k \arccos(x)), \quad k = 0, 1, 2 \ldots \tag{2}$$

Does not really look like a polynomial, doesn't it? Well, it actually is. For $k = 0$ and $k = 1$, the answers are trivial. Indeed $T_0(x) = 1$, $T_1(x) = x$. For the general case, recall the high school formulas:

$$\begin{aligned} \cos(\alpha + \beta) &= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \\ \cos(\alpha - \beta) &= \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) \end{aligned}$$

Combining the two, we obtain:

$$\cos(\alpha + \beta) = 2\cos(\alpha)\cos(\beta) - \cos(\alpha - \beta)$$

Set:

$$\alpha = (k-1)\arccos(x), \quad \beta = \arccos(x)$$

to obtain the recursion:

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), \qquad T_0(x) = 1, \quad T_1(x) = x. \tag{3}$$

We can now use an inductive type of argument to verify that $T_k(x)$ are indeed polynomials. What is so ineresting about Chebyshev polynomials. Well..

many things, but for now we are interested in the roots of the chebyshev polynomials. The roots of $T_n(x)$ are given by

$$x_i = \cos\left(\frac{(2i-1)\pi}{2n}\right), \quad \text{for } i = 1, \dots n$$

The roots are interesting because of the following result.

Theorem 4. *The roots of the Chebyshev polynomial $T_n(x)$ minimize:*

$$\min_{x_1,x_2,\dots,x_n \in [a,b]} \max_{x \in [a,b]} \prod_{k=1}^{n} |x - x_k|$$

Proof. Let $x_i = \cos\left(\frac{(2i-1)\pi}{2k}\right)$ for $i = 1, \dots, n$. Then:

$$\prod_{i=1}^{n}(x - x_i) = \prod_{i=1}^{n}(x - \cos\left(\frac{(2i-1)\pi}{2k}\right)) = \frac{1}{2^{n-1}}T_n(x)$$

Notice that $T_n(\hat{x}_l) = (-1)^l$ at $\hat{x}_l = \cos\left(\frac{l\pi}{n}\right)$ for $l = 0, \dots, n$. These are the $n+1$ locations were T_n attains its extremal values. Our claim is that:

$$\begin{aligned}
\min_{x_1,x_2,\dots,x_n \in [a,b]} \max_{x \in [a,b]} \prod_{k=1}^{n} |x - x_k| &= \prod_{i=1}^{n} \left| x - \cos\left(\frac{(2i-1)\pi}{2n}\right) \right| \\
&= \frac{1}{2^{n-1}} |T_n(x)| \\
&\leq \frac{1}{2^{n-1}}
\end{aligned}$$

We will use contradiction to verify this claim. Assume that there exists another set of nodes for which $\max_{x \in [a,b]} \prod_{k=1}^{n} |x - x_k|$ is minimized. Call the polynomial associated with those nodes $P(x)$. By definition it must be that $|P(x)| < \frac{1}{2^{n-1}}$. Consider the difference between these polynomials:

$$E(x) = \frac{1}{2^{n-1}}T_n(x) - P(x)$$

Since $|P(x)| < \frac{1}{2^{n-1}}$, we have:

$$(-1)^l E(x_l) > 0 \quad \text{for } l = 0, \dots, n$$

This means that $E(x)$ has (at least) n roots. But this is contradiction because degree of $E(x)$ (the leading terms cancel!) $n-1$ (hence it cannot have more than $n-1$ roots). $\square$

It may be surprising to you that the optimal interpolation points are not equispaced. In fact, choosing equispaced samples are a terrible choice. The sequence of polynomials interpolations at the Chebyshev zeros, which have a higher density of points concentrated at the edges, does a better job. As a matter of fact, if the underlying function satisfies some basic "smoothness properties", then convergence is gauranteed.