1 Polynomial Interpolation

Given n distinct points, we can find a unique degree $n-1$ polynomial that passes through these points. Let the polynomial p be

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}.$$

Let the n points be

$$p(x_1) = y_1, p(x_2) = y_2, \cdots, p(x_n) = y_n,$$

where $x_1 \neq x_2 \neq \cdots \neq x_n$.

We can construct a matrix-vector equation as follows to recover the polynomial p.

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix}}_{\vec{a}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\vec{y}}$$

We can solve for the a values by setting:

$$\vec{a} = A^{-1}\vec{y}$$

Note that the matrix A is known as a Vandermonde matrix whose determinant is given by

$$\det(A) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

Since $x_1 \neq x_2 \neq \cdots \neq x_n$, the determinant is non-zero and A is always invertible.

2 Polynomial Regression

Sometimes we may want to fit our data to a polynomial with an order less than $n-1$. If we fit the data to a polynomial of order $m < n$ we get:

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{m-1} x^{m-1}$$

Now when we construct the matrix-vector equation to recover polynomial p, we get:

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{m-1} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{m-1} \end{bmatrix}}_{\vec{a}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ \vdots \\ y_n \end{bmatrix}}_{\vec{y}}$$

With this matrix equation, we have n equations with m unknowns, which means our system is over-defined (since $m < n$). One way to find the best fitting a values for this polynomial is to use least-squares, where you set:

$$\vec{a} = (A^T A)^{-1} A^T \vec{y}$$

3 Lagrange Interpolation

In practice, to approximate some unknown or complex function $f(x)$, we take n evaluations/samples of the function, denoted by $\{(x_i, y_i \triangleq f(x_i));\ 0 \le i \le n-1\}$. For the rest of this question, we will consider the following three points: $\{(0,3),(1,4),(3,-6)\}$.

a) **Using the interpolation method discussed above, find the matrix A such that $A\vec{a} = \vec{y}$.**

Answer

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ -6 \end{pmatrix}$$

b) **What are the coefficients a_0, a_1, a_2?**

Answer

We have our A matrix construction from part (a). All that we have to do now is is invert the matrix to find that

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -4/3 & 3/2 & -1/6 \\ 1/3 & -1/2 & 1/6 \end{pmatrix}$$

Therefore we have $\vec{a} = A^{-1}\vec{y} = \begin{pmatrix} 3 \\ 3 \\ -2 \end{pmatrix}$, which suggests that $f(x) = 3 + 3x - 2x^2$.

c) Observe that this system very quickly becomes frustrating to solve—as n increases, the difficulty of calculating the inverse increases far more quickly.

This is where *Lagrange interpolation* can be useful; the idea of Lagrange interpolation is that, instead of writing the polynomial in question in terms of $\{1, x, x^2\}$, we will write it in terms of $\{L_0(x), L_1(x), L_2(x)\}$, where each

$$L_i(x_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \ne i \end{cases}$$

With that, the problem reduces to finding these new coefficients b_0, b_1, b_2 of the function

$$f(x) = b_0 L_0(x) + b_1 L_1(x) + b_2 L_2(x)$$

such that $f(x_i) = y_i,\ \forall i = 0, 1, 2$. **What are these coefficients b_i?**

Answer

We will show this more generally, and apply our findings to the case wherein $n = 3$. Conveniently by construction, we have

$$f(x_0) = b_0 L_0(x_0) + b_1 L_1(x_0) + \cdots + b_{n-1} L_{n-1}(x_0) = b_0 = y_0$$
$$\vdots$$
$$f(x_{n-1}) = b_0 L_0(x_{n-1}) + b_1 L_1(x_{n-1}) + \cdots + b_{n-1} L_{n-1}(x_{n-1}) = b_{n-1} = y_{n-1}$$

Arranging this in matrix form yields

$$I_n \vec{b} = \vec{y}$$

which imposes $\vec{b} = \vec{y} = \begin{pmatrix} 3 \\ 4 \\ -6 \end{pmatrix}$. No matrix inversion is required!

d) **Show that if we define**

$$L_i(x) = \prod_{j=0; j \neq i}^{n-1} \frac{(x - x_j)}{(x_i - x_j)}$$

then the condition requested from part (c) is satisfied.

Answer

Plugging in x_i, we will get

$$L_i(x_i) = \prod_{j=0; j \neq i}^{j=n-1} \frac{(x_i - x_j)}{(x_i - x_j)} = 1$$

Consider now plugging in x_k into $L_i(x)$, for some $k \neq i$. The term $\frac{x_k - x_k}{x_i - x_k} = 0$ must appear when the summation index $j = k$. Hence $L_i(x_k)$ must be zero for all $k \neq i$.

The intuition is that, since we want some polynomial $L_i(x_j) = 0$ for $j \neq i$, we can take $\prod_{k \neq i}(x - x_k)$. To have $L_i(x_i) = 1$, we can simply "normalize" the polynomial by $\prod_{k \neq i}(x_i - x_k)$, which gives the form of $L_i(x)$.

e) Based on the previous two parts, write down the explicit form of $f(x)$ that passes through the samples $\{(0, 3), (1, 4), (3, -6)\}$ in terms of x as opposed to $L_i(x)$. The resulting formula is the so called Lagrange polynomial which passes through the n sampled points. Does this agree with the previous method?

Answer

Recall that $b_i = y_i$ when using Lagrange polynomials. Using the definition of h:

$$\begin{aligned} f(x) &= b_0 L_0(x) + b_1 L_1(x) + b_2 L_2(x) \\ &= y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) \\ &= 3\left(\frac{(x-1)(x-3)}{(0-1)(0-3)}\right) + 4\left(\frac{(x-0)(x-3)}{(1-0)(1-3)}\right)(-6)\left(\frac{(x-0)(x-1)}{(3-0)(3-1)}\right) \\ &= (x^2 - 4x + 3) + (-2)(x^2 - 3x) + (-1)(x^2 - x) \\ &= -2x^2 + 3x + 3 \end{aligned}$$

Which does agree with our previous method! If we were to test the Lagrange method of determining polynomial coefficients for a much higher-order polynomial (say, $n = 1000$), we would find this method would be several orders of magnitude faster than the one we used in part (b), which is very useful for us to know.

f) Now, suppose instead we wanted to use regression to fit our 3 points to a linear system $f(x) = a_0 + a_1 x$. What are the best-fit coefficients a_0 and a_1 in this situation?

Answer

$$A\vec{a} = \vec{y}$$

$$\begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ 1 & x_2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ -6 \end{pmatrix}$$

From the note, we have $\vec{a} = \left(A^T A\right)^{-1} A^T \vec{y}$. We can find $\left(A^T A\right)^{-1}$:

$$\begin{aligned} \left(A^T A\right)^{-1} &= \left(\begin{pmatrix} 10 & 4 \\ 4 & 3 \end{pmatrix}\right)^{-1} \\ &= \frac{1}{14} \begin{pmatrix} 3 & -4 \\ -4 & 10 \end{pmatrix} \end{aligned}$$

Plugging this all in, we have:

$$\left(A^T A\right)^{-1} A^T \vec{y} = \begin{pmatrix} 3 & -4 \\ -4 & 10 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \\ -6 \end{pmatrix}$$

$$= \frac{1}{14} \begin{pmatrix} 3 & -4 \\ -4 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ -14 \end{pmatrix}$$

$$= \frac{1}{14} \begin{pmatrix} 59 \\ -144 \end{pmatrix}$$