CS323 LECTURE NOTES - LECTURE 15

1 Least squares

The approximations that we have talked about before find a polynomial that goes exactly through the given points. A least squares approximation can be used to find a polynomial with degree much smaller than the given number of points, such that it passes as close as possible to the points but is not required to go through them.

First we have to define what we mean by "passes as close as possible", to do this we need a measure of error, as well as a formal definition of the problem that we want to solve.
Problem.

Given a set of m points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ and a n-degree polynomial

$$P(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = \sum_{i=0}^{n} a_j x^j$$

We can compute the approximation error for a point (x_i, y_i) as the vertical distance from the point to the polynomial when $x = x_i$:

$$\text{error}_i = |P(x_i) - y_i|$$

so the total error when all the points have been taken in to account is:

$$E_1(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{m} |P(x_i) - y_i|$$

This error is a function of the coefficients $a_0, a_1, \ldots, a_m$ since all other values are constant.

Our problem consists in finding the coefficients $a_0, a_1, \ldots, a_m$ that minimize the total error. For this purpose, we would like to compute the derivatives of E_1 an make them equal to 0, but the absolute value function is not differentiable in 0, so we prefer to use the function x^2 instead of $|x|$, so the error is:

$$E(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{m} (P(x_i) - y_i)^2$$

or

$$E(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{m} (\sum_{i=0}^{n} a_j x^j - y_i)^2$$

Using this formula we can find its minimum using all of its partial derivatives with respect to a_k and making each one equal to 0.

$$\frac{\partial E}{\partial a_k} = \sum_{i=1}^{m} 2(\sum_{j=0}^{n} a_j x_i^j - y_j)(x_i^k) = 0$$

since

$$\frac{\partial a_j}{\partial a_k} = \begin{cases} 1 & \text{si} \quad j = k \\ 0 & \text{si} \quad j \neq k \end{cases}$$

and x_i, y_i are constants. The factor of 2 can be placed outside of the summation where it will be cancelled out by the 0, and the term x_i^k can enter the summation since it is constant:

$$\sum_{i=1}^{m} (\sum_{j=0}^{n} a_j x_i^{k+j} - x_i^k y_i) = 0$$

therefore,

$$\sum_{i=1}^{m} \sum_{j=0}^{n} a_j x_i^{k+k} - \sum_{i=1}^{m} x_i^k y_i = 0$$

We can now swap the summations:

$$\sum_{j=0}^{n} a_j \sum_{i=1}^{m} x_i^{k+j} = \sum_{i=1}^{m} x_i^k y_i \quad \forall k = 0, \ldots, n$$

recall that we have one equation for each value of k:
$k = 0$:

$$m a_0 + (\sum_{i=1}^{m} x_i) a_1 + (\sum_{i=1}^{m} x_i^2) a_2 + \ldots + (\sum_{i=1}^{m} x_i^n) a_n = \sum_{i=1}^{m} y_i$$

$k = 1$:

$$(\sum_{i=1}^{m} x_i) a_0 + (\sum_{i=1}^{m} x_i^2) a_1 + (\sum_{i=1}^{m} x_i^3) a_2 + \ldots + (\sum_{i=1}^{m} x_i^{n+1}) a_n = \sum_{i=1}^{m} x_i y_i$$

$$\vdots$$

$k = n$:

$$(\sum_{i=1}^{m} x_i^n)a_0 + (\sum_{i=1}^{m} x_i^{n+1})a_1 + (\sum_{i=1}^{m} x_i^{n+2})a_2 + \ldots + (\sum_{i=1}^{m} x_i^{2n})a_n = \sum_{i=1}^{m} x_i^n y_i$$

These equations represent a linear system with $n+1$ equations and $n+1$ unknowns which can be written in matrix form:

$$\begin{pmatrix} m & \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \cdots & \sum_{i=1}^{m} x_i^n \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \cdots & \sum_{i=1}^{m} x_i^{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{m} x_i^n & \sum_{i=1}^{m} x_i^{n+1} & \sum_{i=1}^{m} x_i^{n+2} & \cdots & \sum_{i=1}^{m} x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ \vdots \\ \sum_{i=1}^{m} x_i^n y_i \end{pmatrix}$$

Example:

In the case finding a linear approximation $n = 1$, to a collection of points, the system of equations is:

$$\begin{pmatrix} m & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

This system has 2 equations and 2 unknowns and can easily be solved by Cramer's Rule.

$$\begin{aligned} a_0 &= \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{m \sum x_i^2 - (\sum x_i)^2} \\ a_1 &= \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - (\sum x_i)^2} \end{aligned}$$

Analysis

1. The algorithm computes $x_i, x_i^2, \ldots, x_i^{2n}$, that can be placed in a $2n \times m$ matrix, and requires time $O(nm)$ time to be computed.
2. To compute all summations we need time in $O(nm)$
3. The solution of the linear system by Gaussian Elimination requires $O(n^3)$ time.

Therefore, the time complexity of the least squares polynomial approximation algorithm is $O(nm + n^3)$, and since $n \ll m$, the algorithm is in $O(nm)$

1.1 Applications of the least squares linear approximation

The least squares method can also be used to approximate non-linear functions:

1. $y = a\mathrm{e}^{bx}$
2. $y = ax^b$

In the first case we can take natural log on both sides:

$$\ln y = \ln a + bx$$

So we have

$$Y = A + bx$$

Where $Y = \ln y$.

In this case we can use the pairs $(x_i, \ln y_i)$ and apply the least squares method to find the line $Y = a_0 + a_1 x$ that best approximates the given pairs.

Finally we have that $b = a_1$ and $\ln a = a_0$, therefore $a = \mathrm{e}^{a_0}$

In the second case we can take natural log on both sides:

$$\ln y = \ln a + b \ln x$$

So we have

$$Y = A + bX$$

where $Y = \ln y$, and $X = \ln x$.

In this case we can use the pairs $(\ln x_i, \ln y_i)$ and apply the least squares method to find the line $Y = a_0 + a_1 X$ that best approximates the given pairs.

Finally we have that $b = a_1$ and $\ln a = a_0$, therefore $a = \mathrm{e}^{a_0}$