Function Approximation [Curve Fitting]

Two types of problems

a. Data have scatter – Regression

b. Data are precise – Interpolation

Approximation of a complex function by a simpler function

What functions to use for approximation

Polynomials – easy to determine (fit)
estimate
integrate
differentiate

Two more qualities of polynomials

1. Uniform approximation [Stone-Weierstrass approx.]

Any continuous function defined on a closed interval $[a, b]$ can be approximated as closely as desired by a polynomial

$$\left| f(x) - P_n(x) \right| < \varepsilon \qquad a \leq x \leq b$$

$\varepsilon > 0$ can be arbitrarily small

2. Uniqueness A polynomial of degree 'n' passing exactly through $(n+1)$ discrete points is unique

→ Polynomial may be represented in different forms but all forms are equivalent

Regression :- Many approaches

(x_i, y_i) $\qquad \hat{y}_i$ $\qquad i = 1, 2, \ldots N$

$$e_i = y_i - \hat{y}_i$$

(a) Principle of least squares

Minimizes sum of squared errors

$$\min \sum (e_i)^2 \qquad l_2^2 \text{-norm}$$

b) Minimizes sum of absolute errors

$$\min \sum |e_i| \qquad l_1 \text{-norm}$$

c) Minimizes maximum error

$$\min e_{max} \rightarrow l_\infty \text{-norm}$$

Linear Regression

$$\hat{y}_i = a_0 + a_1 x_i$$

$$e_i = (y_i - \hat{y}_i)$$

Sum of squared errors

$$S_{(a_0, a_1)} = \sum (y_i - \hat{y}_i)^2$$

$$\left.\begin{aligned} \frac{\partial S}{\partial a_0} &= 0 \\ \frac{\partial S}{\partial a_1} &= 0 \end{aligned}\right\} \rightarrow \text{Normal Equation}$$

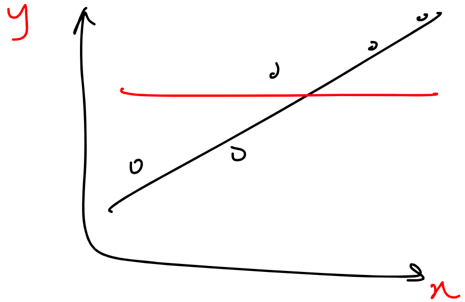
$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Non-linear cases $\Rightarrow$ Linearize the problem

How good is the fit ?

(x_i, y_i)

$\rightarrow \hat{y}_i = a_0 + a_1 x_i$



$$S = \sum e_i^2$$

smaller the value of 's' better is the fit.

Normalize the measure S by a simple model S_0

$$\hat{y}_i = \overline{y} = \frac{1}{N} \sum y_i$$

$$S_0 = \sum (y_i - \overline{y})^2$$

A measure of goodness of fit

$$R^2 = 1 - \frac{S}{S_0} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}$$

R^2 - coefficient of determination (index)

$$0 \leq R^2 \leq 1$$

$R^2 \approx 1$ — Very good fit

$R^2 \approx 0$ $\rightarrow$ poor fit

Example

x	0	2	4	8
y	1	0.7937	0.63	0.3968

$\bar{y} = 0.7051$

$$S_0 = \sum (y_i - \bar{y})^2 = 0.1955$$

$$S_1 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a_0 - a_1 x_i)^2$$

$$= 0.0039$$

$$R^2 = 1 - \frac{S_1}{S_0} = 1 - 0.0197$$

$$= 0.9803$$

Extending the least square fit to higher order polynomials

$$P_1(x) = a_0 + a_1 x$$

$$P_2(x) = a_0 + a_1 x + a_2 x^2$$

Linear least squares

Ordinary least squares

(x_i, y_i) $\quad e_i = (y_i - \hat{y}_i)$

$$S = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$$

$$S(a_0, a_1, a_2) = \sum (y_i - a_0 - a_1 x - a_2 x^2)$$

$$\frac{\partial S}{\partial a_0} = 0 \qquad \frac{\partial S}{\partial a_1} = 0 \qquad \frac{\partial S}{\partial a_2} = 0$$

$$\frac{\partial S}{\partial a_0} = 2\sum \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)(-1)$$

$$\frac{\partial S}{\partial a_1} = 2\sum \left(y_i - \underline{a_0} - a_1 x_i - a_2 x_i^2\right)(-\underline{x_i})$$

$$\frac{\partial S}{\partial a_2} = 2\sum \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)(-x_i^2)$$

$$\begin{bmatrix} N & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

$$P_n(x_i) = a_0 + a_1 x_i + \dots + a_n x_i^n = \sum_{j=0}^{n} a_j x_i^j$$

Normal equation

$$S = \sum e_i^2 = \sum \left[y_i - \left(\sum_{j=0}^{n} a_j x_i^j\right)\right]^2$$

$$\frac{\partial S}{\partial a_j} = 0 \qquad j = 0, 1, \dots n$$

$$\begin{bmatrix} N & \sum x_i & \sum x_i^2 & \dots & \sum x_i^n \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^{n+1} \\ \vdots & & & & \\ \sum x_i^n & \dots & \dots & \dots & \sum x_i^{n+n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \\ \vdots \\ \sum y_i x_i^n \end{bmatrix}$$

Extending Ordinary least squares to a general basis function

Poly

$$\hat{y}_i = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n$$

Any function ϕ

$$\hat{y}_i = a_0 \phi_0(x) + a_1 \phi_1(x) + a_2 \phi_2(x) + \cdots + a_n \phi_n(x)$$

$\phi_0(x) = 1$

$\phi_1(x) = x$

$\phi_2(x) = x^2$

$\vdots$

$\phi_n(x) = x^n$

ϕ's are the basis functions

Design Matrix

(x_i, y_i) $i = 1, 2, \ldots N$

Matrix Form

n - basis functions

$$\underbrace{\begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \cdots & \phi_n(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \cdots & \phi_n(x_2) \\ \vdots & & & \\ \phi_0(x_N) & \phi_1(x_N) & \cdots & \phi_n(x_N) \end{bmatrix}}_{\Phi} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

Φ - Design matrix

Quadratic Case

$$\hat{y}_i = a_0 + a_1 x_i + a_2 x_i^2$$

Design matrix

$$\bar{\Phi} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}_{N \times 3} \qquad a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

Normal equation

$$\boxed{\bar{\Phi}^T \bar{\Phi}\, a = \bar{\Phi}^T y}$$

Problem $\rightarrow$ $\bar{\Phi}^T \bar{\Phi}$ has a large condition number

(a) Better way to solve normal equation

Singular value decomposition

(b) Select basis function (ϕ_i) so that the condition number of $\bar{\Phi}^T \bar{\Phi}$ can be reduced.

$\Rightarrow$ One option is to take basis functions which are orthogonal, so that $\bar{\Phi}^T \bar{\Phi}$ will be a diagonal matrix

Orthogonal Basis Functions

Two vectors $X = [x_1\ x_2\ \dots\ x_n]^T$ and

$Y = [y_1\ y_2\ \dots\ y_n]^T$ are said to

be orthogonal

$$X^T Y = 0$$

$$\Rightarrow \quad \Rightarrow \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = 0$$

In 2 and 3 dimensions, it means that the vectors are perpendicular

It is customary to say, that the vectors are perpendicular to each other in n dimensional space

Assume, that the number of dimensions increase to infinity

Vectors — continuous function

Summation — integral

$$\langle x, y \rangle = \int_a^b x(t)\, y(t)\, dt = 0$$

Then function $x(t)$ and $y(t)$ in the range $[a, b]$ are called orthogonal functions

Infinite dimensional space → Hilbert space

For convenience, sometimes a weight

$W(t) > 0$ is introduced

$$\int_a^b W(t)\, x(t)\, y(t)\, dt = 0$$

Examples $\cos x, \sin x$ 0, to 2π

Orthogonal Polynomials