Least Squares Optimization

Monday, August 15, 2022 2:55 PM

Suppose we have points (x_n, y_n) for $n \in [1, \ldots, N]$.
How can we fit a polynomial to this set of points?

$\boxed{\text{Fitting a Line}}$

We have a line $y = mx + b$. One way to select ideal coefficients m and b is to minimize the square error btw $mx+b$ and the data. We seek m and b as follows:

$$\min_{m,b} E^2(m,b) = \min_{m,b} \sum_{n=1}^{N} \left[y_n - (mx_n + b) \right]^2$$

We can best organize these ideas with matrices.

$$X \equiv \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Let $\vec{a} = \begin{bmatrix} b \\ m \end{bmatrix}$

$\vec{y} - X\vec{a}$ is a vector of unsquared errors.

$(\vec{y} - X\vec{a})^T (\vec{y} - X\vec{a})$ squares and sums these errors using the dot product.

This quantity is minimized when

$$\nabla_{\vec a}\left[(\vec y - X\vec a)^T(\vec y - X\vec a)\right] =$$

$$\nabla_{\vec a}\left[\vec y^T\vec y - \vec a^T X^T\vec y - \vec y^T X\vec a + \vec a^T X^T X\vec a\right]$$

$\vec 0$ vector is sometimes just written 0

$$\nabla_{\vec a}\left[\vec y^T\vec y\right] = \vec 0 = \begin{bmatrix}0\\0\end{bmatrix}$$

$$\nabla_{\vec a}\left[-\vec a^T X^T\vec y\right] = \nabla_{\begin{bmatrix}b\\m\end{bmatrix}}\left[-[b\ m]\underbrace{X^T}_{2\times N}\underbrace{\vec y}_{N\times 1}\right]$$

$$= -X^T\vec y.$$

$$\nabla_{\vec a}\left[-y^T\vec X\vec a\right] = \nabla_{\vec a}\left[-\vec a^T\ \vec X^T\ \vec y\right] = -X^T\vec y \quad \text{shown above.}$$

$$\nabla_{\vec a}\left[\vec a^T X^T X\vec a\right] = \nabla_{\vec a}\left[(\underbrace{X}_{N\times 2}\underbrace{\vec a}_{2\times 1})^T X\vec a\right]$$

<u>Note:</u>

$\vec a^T M\vec a$ where $M = M^T$ is known as a quadratic form.

what is $\nabla_{\vec a}\,\vec a^T M\vec a$?

$$M\vec a = \begin{bmatrix} M_{11}a_1 + M_{12}a_2 + \cdots + M_{1N}a_N \\ M_{21}a_1 + M_{22}a_2 + \cdots + M_{2N}a_N \\ \vdots \end{bmatrix}$$

$$\vec a^T M\vec a = M_{11}a_1^2 + M_{12}a_1a_2 + \cdots + M_{1N}a_1a_N$$
$$+ M_{21}a_1a_2 + M_{22}a_2^2 + \cdots + M_{2N}a_2a_N + \cdots$$
$$= \sum_{m=1}^{N}\sum_{n=1}^{N} M_{mn}a_ma_n$$

Observe that

$$\frac{\partial}{\partial a_k}\sum_{m=1}^{N}\sum_{n=1}^{N} M_{mn}a_ma_n =$$

$$\frac{\partial}{\partial a_k} \sum_{m=1} \sum_{n=1} M_{mn} a_m a_n =$$

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \frac{\partial}{\partial a_k} M_{mn} a_m a_n =$$

$$2 M_{kk} a_k + \sum_{m \neq k} M_{mk} a_m + \sum_{n \neq k} M_{kn} a_n =$$

$$2 \sum_{n=1}^{N} M_{kn} a_n \quad \text{by symmetry of } M.$$

$$2 \cdot \begin{array}{c} k^{th} \\ \text{row of} \\ M \end{array} \begin{bmatrix} \cdots \\ \\ \\ \end{bmatrix} \underbrace{\begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix}}_{\vec{a}}$$

$\downarrow$ looks like

$$\nabla_{\vec{a}}\, \vec{a}^T M \vec{a} = 2M\vec{a}$$

End note.

$$\nabla_{\vec{a}} \left(\vec{a}^T X^T X \vec{a} \right) = 2 X^T X \vec{a}$$

So we arrive at $\nabla_{\vec{a}} \left(\|\vec{y} - X\vec{a}\|^2 \right)$ $= -2X^T\vec{y} + 2X^T X\vec{a} = 0$ as our optimization equation.

$$X^T \vec{y} = X^T X \vec{a}$$

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

is the optimal solution for coefficients.

$X^T X$ is invertible if we have 2 or more unique points in our data.

$X^T X$ is invertible if we have 2 or more unique points in our data.

Fitting a Polynomial

Fit (x_i, y_i) to $ax^2 + bx + c = \hat{y}$

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

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

$$\vec{a}^* = (X^T X)^{-1} X^T \vec{y}$$ as above.

We can arbitrarily generalize to any degree. But what degree is appropriate? The one that minimizes the "minimum description length criterion":

$$mdl(p) = \ln\left(\epsilon^2(P)\right) + \frac{p \ln(N)}{N}$$

Fitting a Sinusoid

How do we fit $s[n]$ of length N to a sinusoid $y = A\cos(2\pi f_0 n + \phi)$?

$$\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$$

$$y = A\cos(2\pi f_0 n)\cos(\phi) - A\sin(2\pi f_0 n)\sin(\phi)$$

$$a_1 \equiv A\cos(\phi) \qquad a_2 \equiv A\sin(\phi)$$

$$A = \sqrt{a_1^2 + a_2^2} \qquad \phi = \tan^{-1}\left(\frac{a_2}{a_1}\right)$$

$$A = \sqrt{a_1^2 + a_2^2} \qquad \phi = \tan^{-1}\left(\frac{a_2}{a_1}\right)$$

$$\vec{a} \equiv \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 0 \\ \cos(2\pi f_0) & \sin(2\pi f_0) \\ \cos(4\pi f_0) & \sin(4\pi f_0) \\ \vdots & \vdots \\ \cos(2\pi f_0(N-1)) & \sin(2\pi f_0(N-1)) \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\vec{a}} = \underbrace{\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix}}_{\vec{g}}$$

$$\vec{a}^* = (X^T X)^{-1} X^T \vec{g}$$

$$\begin{bmatrix} 1 & \cos(2\pi f_0) & \cdots & \cos(2\pi f_0(N-1)) \\ 0 & \sin(2\pi f_0) & \cdots & \sin(2\pi f_0(N-1)) \end{bmatrix} \begin{array}{l} \leftarrow \vec{c}^T \\ \leftarrow \vec{s}^T \end{array}$$

$$\begin{array}{cc} \vec{c} & \vec{s} \end{array}$$

$$\begin{bmatrix} 1 & 0 \\ \cos(2\pi f_0) & \sin(2\pi f_0) \\ \vdots & \vdots \\ \cos(2\pi f_0(N-1)) & \sin(2\pi f_0(N-1)) \end{bmatrix}$$

$$= X^T X = \begin{bmatrix} \vec{c}^T\vec{c} & \vec{c}^T\vec{s} \\ \vec{s}^T\vec{c} & \vec{s}^T\vec{s} \end{bmatrix}$$

Assume that $f_0 = \frac{K}{N}$, a harmonic of the signal's length.

$$\vec{c}^T\vec{c} = 1 + \cos^2(2\pi f_0) + \cdots + \cos^2(2\pi f_0(N-1))$$

$$= 1 + \frac{1}{2}(1 + \cos(4\pi f_0)) + \cdots + \frac{1}{2}(1 + \cos(4\pi f_0(N-1)))$$

$= N/2$ since the sinusoid parts cancel over an entire period.

$\vec{c}^T\vec{s} = \vec{s}^T\vec{c} = 0$ since $\vec{s}$ and $\vec{c}$ are orthogonal.

entire period.

$\vec{c}^T \vec{s} = \vec{s}^T \vec{c} = 0$ since $\vec{s}$ and $\vec{c}$ are orthogonal.

$$\vec{s}^T \vec{s} = 1 + \frac{1}{2}(1-\cos(4\pi f_0)) + \cdots + \frac{1}{2}\left(1-\cos(4\pi f_0(N-1))\right)$$
$$= N/2$$

So $X^T X = \begin{bmatrix} N/2 & 0 \\ 0 & N/2 \end{bmatrix}$

and $(X^T X)^{-1} = \begin{bmatrix} 2/N & 0 \\ 0 & 2/N \end{bmatrix}$

$$\boxed{\frac{2}{N} X^T \vec{y} = \vec{a}^*}$$