Tutorial - Week 8

Note Title 03/03/2015

Normal equation for linear regression: (a.k.a standard form for least square problem)

Suppose we have an overdetermined system: $Ax = b$ with $A_{m \times n}$ $(m > n)$, $x_{n \times 1}$ and $b_{m \times 1}$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix}$$

No exact solution

If $S(x)$ is the set with all the possible solutions we want to find $\hat{x} \in S(x)$ such that:

$$\hat{x} = \underbrace{\operatorname{argmin}}_{\text{solution with the minimum square error}} (S(x))$$

We can rewrite our system as $Ax - b = 0$. Since $S(x)$ is a set of solutions we can try to minimize it to find $\hat{x}$.

To do that in a least square sense, we can write the 2-norm for our system, then set the derivative of it to $\emptyset$:

$$\|S(x)\|_2 = \sum_{i=1}^{m} \left| \left(\sum_{j=1}^{n} A_{ij} x_j \right) - b_i \right|^2 \equiv \|Ax - b\|^2 = (A^T x^T - b^T)(Ax - b)$$

we want to take the derivative with respect to the vector x of this and then set it to zero

$$\frac{d}{dx_j} \underline{(A^T x^T - b^T)(Ax - b)} = 0$$

partial derivative with respect to each component of $\bar{x}$

$$\frac{d}{dx_j} \Big(\underbrace{A^T x^T A x - A^T x^T b - b^T A x + b^T b}_{\text{expanded the products}} \Big) = 0$$

$A^T x^T b = b^T A x$ (column vector, same dimension of b)

$$\frac{d}{dx_j}\left(A^T x^T A x - 2A^T x^T b + \cancel{b^T b}\right) = 0 \iff 2A^TA\hat{x} - 2A^Tb = 0$$

$$2A^TAx \qquad -2A^Tb \qquad 0$$

$$A^TA\hat{x} - A^Tb = 0 \iff A^TAx = A^Tb$$

$$\boxed{\hat{x} = (A^TA)^{-1}A^Tb}$$ normal eq. for linear regression

Numerical example:

$Ax = b$ where:

$$\underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix}}_{b}$$

$A_{m\times n}$ $m > n$

$x_{n\times 1}$

$b_{m\times 1}$

To compute $\hat{x}$ I need A^T:

$$A^T = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}$$

$$A^T A x = A^T b$$

$$\begin{bmatrix} 1 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix} \Longleftrightarrow \begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 22 \\ 10 \end{pmatrix}$$

I can solve this using Gaussian elimination:

$$\left(\begin{array}{cc|c} 6 & 2 & 22 \\ 2 & 3 & 10 \end{array}\right) \quad R_2 = -3R_2 + R_1 \quad \left(\begin{array}{cc|c} 6 & 2 & 22 \\ 0 & -7 & -8 \end{array}\right)$$

$$\begin{pmatrix} 6 & 2 \\ 0 & -7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 22 \\ -8 \end{pmatrix} \rightarrow \begin{matrix} x_2 = 8/7 \\ x_1 = 23/7 \end{matrix} \rightarrow \hat{x} = \begin{pmatrix} 23/7 \\ 8/7 \end{pmatrix}$$

Taylor polynomials

If $f(x)$ is a function, and $f^{(k)}(x)$ is continuous in $[a, b]$ and $f^{(k+1)}(x)$ exists, the the k^{th} Taylor polynomial for $f(x)$ is:

$$T_k(x) = \sum_{i=0}^{k} \frac{f^{(k)}(c)}{k!} (x-c)^k$$

c is where I evaluate my polynomial.

If $c = \emptyset \rightarrow$ Mclaurin: $T_k(x) = \sum_{i=0}^{k} \frac{f^{(k)}(0)}{k!} \cdot x^2$

Example: consider $f(x) = \cos(x)$ at $c = 0$

the Mclaurin series for $f(x)$ is $\sum_{i=0}^{\infty} \frac{(-1)^k}{2k!} x^{2k}$

So the Taylor polynomials, at $c=0$, are defined as:

$$T_k(x) = \sum_{i=0}^{k} \frac{(-1)^k}{2k!} x^{2k}$$

$k=0 \rightarrow T_0(x) = \frac{(-1)^0}{0!} x^0 = 1$

$k=1 \rightarrow T_1(x) = \frac{(-1)^0}{0!} x^0 + \frac{(-1)^1}{2!} x^2 = 1 - \frac{1}{2} x^2$

$k=2 \rightarrow T_2(x) = \frac{(-1)^0}{0!} x^0 + \frac{(-1)^1}{2!} x^2 + \frac{(-1)^2}{4!} x^4 = 1 - \frac{1}{2} x^2 + \frac{1}{24} x^4$

See .m files in the directory for a good idea of how this polynomials behave! :)