Some useful linear algebra & linear regression

Zhenzhen Kou

Thursday, January 20, 2005

Some useful linear algebra

- Scalar, Vector, Matrix
- Basic operations (+, -, *)
- Dot products (Inner product), Length of a vector:|| x ||
- Transpose, Inverse
- Matrix calculus

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^{\mathbf{T}}\mathbf{A}) = \mathbf{A}\,, \frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}^{\mathbf{T}}\,, \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^{\mathbf{T}}\mathbf{x}) = 2\mathbf{x}$$

The chain rule: $\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{z}}{\partial \mathbf{y}}$,

The regression problem

- **Instances:** $\langle \mathbf{x}_j, t_j \rangle$
- **Learn:** Mapping from x to t(**x**)
  - Find coeffs **w**={$w_1,\ldots,w_k$} in

$$\underbrace{t(\mathbf{x})}_{\text{data}} \approx \widehat{f}(\mathbf{x}) = \sum_i w_i h_i(\mathbf{x})$$

- Precisely, minimize the sum of squares residua (SSR):

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_j \left(t(\mathbf{x}_j) - \sum_i w_i h_i(\mathbf{x}_j) \right)^2$$

Why minimize SSR?

- Learned in class: maximizing log-likelihood in a Gaussian model

$$P(t \mid \mathbf{x}, \mathbf{w}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left[t - \sum_i w_i h_i(\mathbf{x})\right]^2}{2\sigma^2}}$$

- Minimizing SSR is also called Least Square Error

Linear regression

- Problem: Find coeffs **w**={w_1,…,w_k} in

$$\underbrace{t(\mathbf{x})}_{\text{data}} \approx \hat{f}(\mathbf{x}) = \sum_i w_i h_i(\mathbf{x})$$

- Start from linear regression

$$\mathbf{y} = t(\mathbf{x}) = \sum_i w_i x_i = \mathbf{w}^{\mathbf{T}}\mathbf{x}$$

- A trick to deal with the constant term:

$$\mathbf{x} = (x_1, ..., x_{m-1}, 1)^T$$

Linear regression

Dataset has form

x_1	y_1
x_2	y_2
x_3	y_3
.:	:
.	
x_R	y_R

Write matrix X and Y thus:

$$\mathbf{x} = \begin{bmatrix} \ldots\mathbf{x}_1\ldots \\ \ldots\mathbf{x}_2\ldots \\ \vdots \\ \ldots\mathbf{x}_R\ldots \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ & & \vdots & \\ x_{R1} & x_{R2} & \ldots & x_{Rm} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_R \end{bmatrix}$$

Each row of X represents a record

Let **Y** represent the output data and **X** represent the input data, the linear regression model assumes a vector **w** such that

$$Y = X \bullet \mathbf{w} \qquad \text{Here } w = \begin{bmatrix} w_1 \\ w_1 \\ \ldots \\ w_m \end{bmatrix}$$

Linear regression

•To find the maximum likelihood estimation of $\mathbf{w}$ is to look for $\mathbf{w}$ which minimizes $\left\|\mathbf{Xw} - \mathbf{Y}\right\| = (\mathbf{Xw} - \mathbf{Y})^{\mathbf{T}}(\mathbf{Xw} - \mathbf{Y})$

a concave function of $\mathbf{w}$

•Solve $\dfrac{\partial(\mathbf{Xw} - \mathbf{Y})^{\mathbf{T}}(\mathbf{Xw} - \mathbf{Y})}{\partial \mathbf{w}} = \mathbf{0}$

Aplly the chain rule, we get: $\dfrac{\partial(\mathbf{Xw} - \mathbf{Y})^{\mathbf{T}}}{\partial \mathbf{w}} \cdot 2(\mathbf{Xw} - \mathbf{Y}) = \mathbf{0}$

i.e. $\mathbf{X}^{\mathbf{T}}(\mathbf{Xw} - \mathbf{Y}) = \mathbf{0}$

• Therefore $\mathbf{w} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$

What if X^TX does not have inverse?
Pseudoinverse!

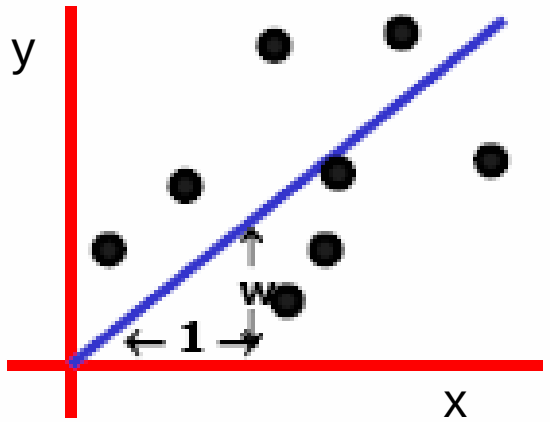
Linear regression

The max. likelihood $\boldsymbol{w}$ is $\boldsymbol{w} = (X^T X)^{-1}(X^T Y)$

$X^T X$ is an m x m matrix: i,j'th elt is $\sum_{k=1}^{R} x_{ki} x_{kj}$

$X^T Y$ is an m-element vector: ith elt $\sum_{k=1}^{R} x_{ki} y_k$

Linear Regression



DATASET

inputs	outputs
$x_1 = 1$	$y_1 = 1$
$x_2 = 3$	$y_2 = 2.2$
$x_3 = 2$	$y_3 = 2$
$x_4 = 1.5$	$y_4 = 1.9$
$x_5 = 4$	$y_5 = 3.1$

The regression problem
What if non-linear

- **Learn:** Mapping from x to t(**x**)
  - Find coeffs **w**={w_1,…,w_k} in

$$\underbrace{t(\mathbf{x})}_{\text{data}} \approx \widehat{f}(\mathbf{x}) = \sum_i w_i h_i(\mathbf{x})$$

- What if h(x) is a non-linear function?
  - Logistic regression

What you need to know

- Linear regression
  - Optimizing sum squared error
  - A linear regression has a close form solution
- More about regression
  - Andrew Moore's tutorials at
    http://www-2.cs.cmu.edu/~awm/tutorials/

Questions?