宋绍铭 自24 2012011467

1.a 对于最小二乘法，需要使得

$$J(\theta)=\frac{1}{2}\sum_{i=1}^{m}(y^i-h_\theta(x^i))^2$$

最小，因此对于

$$J(\theta)=\frac{1}{2}(X\theta-\vec{y})^T(X\theta-\vec{y})$$

求极值，我们得到

$$\nabla J(\theta)=\nabla_\theta\frac{1}{2}(X\theta-\vec{y})^T(X\theta-\vec{y})$$

$$=X^TX\theta-X^T\vec{y}=0\Rightarrow\theta=(X^TX)^{-1}X^T\vec{y}$$

而对于极大似然估计，我们令 $L(y:\varphi)=\prod_{i=1}^{T}P(y_i)$ 有极值，得到

$$\frac{\partial}{\partial\varphi_i}L(y:\varphi)=0$$

对于题中给出的 $y=x^t\beta+\varepsilon$ 的形式

有 $\ln L(\beta,\sigma^2)=-\frac{T}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu_t)^2$

$$=-\frac{T}{2}\ln(2\pi\sigma^2)-\frac{1}{\sigma^2}\cdot J(\theta)$$

~~其中~~，

~~因此~~ $\because -\frac{T}{2}\ln(2\pi\sigma^2)$ 为常数

$\therefore J(\theta)$ 取到极值时，$L(\beta,\sigma^2)$ 也取到极值.

$\therefore$ 这两种方法在求解 β 时是等价的.

1.b 对于单变量回归，由1.a中的结果可以有

$$\beta_i^*=(x_i^Tx_i)^{-1}x_i^T\vec{y}=\frac{x_i^Ty}{x_i^Tx_i}$$

对于多变量回归，$\beta^*=[\beta_1^*,\beta_2^*\cdots]^T$

若令 $\beta=\beta^*$，则必有 $(X^TX)^{-1}=\begin{bmatrix}\frac{1}{x_1^Tx_1} & & \\ & \frac{1}{x_2^Tx_2} & \\ & & \ddots\end{bmatrix}$

显而易见，当 $x_1,x_2,\cdots,x_n$ 之间不存在严格相关时有 $\beta_i=\beta_i^*$，

否则 $\beta_i\neq\beta_i^*$