线性回归

$$Y = f(X) + \epsilon$$

input

x_1^m
x_2^m
x_3^m

output

y^m
$Y = M\beta \Rightarrow M^TY = M^TM\beta \Rightarrow \beta = (M^TM)^{-1}M^{-1}Y$

$$y = ax_1 + bx_2 + cx_3 + d + \epsilon$$

非线性回归

$$y = ax_1^2 + b\log x_2 + ce^{x_3} + d$$

input

$$\begin{pmatrix} x_1^2 \\ \log x_2 \\ e^{x_3} \\ 1 \end{pmatrix}$$

$$\min_\beta \quad \|X\beta - Y\|^2 + \lambda\|\beta\|^2$$

凸函数

$$f(\lambda\beta_1 + (1-\lambda)\beta_2) \geq \lambda f(\beta_1) + (1-\lambda)f(\beta_2)$$

伪梯度

$y = |x_1| + |x_2| + |x_3|$
$g = \operatorname{sgn}(x_1, x_2, x_3)$

求解

$$\min_\beta \quad \|X\beta - Y\|_2^2 + \lambda\|\beta\|_2^2$$

- 求导

$$\frac{\partial(X\beta - Y)^T(X\beta - Y) + \lambda\beta^T\beta}{\partial\beta}$$

$$(X\beta - Y)^TX + \lambda\beta^T = 0$$

$$(X^TX + \lambda I)^T\beta = X^TY$$

矩阵乘法

$$C = A \cdot B \Rightarrow C_{ij} = \sum_k A_{ik}B_{kj}$$

$$D = A \cdot B \cdot C \Rightarrow D_{ij} = \sum_{k_1} \sum_{k_2} A_{ik_1} B_{k_1 k_2} C_{k_2 j}$$

矩阵求导

$$f(X) = AX + b$$

- 变量是向量，函数是向量

$$\frac{\partial(f_1, \ldots, f_m)}{\partial(x_1, \ldots, x_n)}$$

$$D_{ij} = \frac{\partial f_i}{\partial x_j} = A_{ij}$$

  则

$$\frac{\partial f}{\partial X} = A$$

  A是矩阵，x是实数$A = f(x)$
  则

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1}$$

- 变量是矩阵，函数是标量

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{n1}} & \cdots & \frac{\partial f}{\partial x_{nn}} \end{bmatrix}$$

- 迹
  迹
  `tr` 是矩阵对角线元素之和

$$\frac{\partial tr[AXB]}{\partial x}$$

$$[AXB]_{hk} = \sum_i \sum_j A_{hi} X_{ij} B_{jk}$$

  A是个矩阵
  A的迹 $tr[A] = \sum_{i=1}^n a_{ii}$

$$tr[AXB] = \sum_h \sum_i \sum_j A_{hi} X_{ij} B_{jh}$$

$$\because \quad \frac{\partial tr[AXB]}{\partial x_{ij}} = \sum_h A_{hi} B_{jh} = \sum_h B_{jh} A_{hi} = [BA]_{ji} = [(BA)^T]_{ij}$$

$$\therefore \quad \frac{\partial tr[AXB]}{\partial x} = A^T B^T$$

$$\frac{\partial tr[AX^TB]}{\partial x}=\frac{\partial tr[B^TXA^T]}{\partial x}=BA$$

$$\frac{\partial tr[AXBXC]}{\partial x}=A^TC^TX^TB^T+B^TX^TA^TC^T$$

- 范式

$$\|A\|_F^2=\sum_i\sum_j a_{ij}^2$$

$$[AA^T]_{ii}=\sum_k A_{ik}^2$$

$$tr[AA^T]=\sum_i\sum_j A_{ij}^2=\|A\|_F^2$$

$$\frac{\partial\|B-XA\|_F^2}{\partial x}=\frac{\partial tr[(B-XA)^T(B-XA)]}{\partial x}$$

$$=\frac{\partial tr[B^TB+A^TX^TXA-A^TX^TB-B^TXA]}{\partial x}$$

$$=2XAA^T-2BA^T$$

令$2XAA^T-2BA^T=0$得:

$$X=BA^T(AA^T)^{-1}$$

logistic regression $\Rightarrow$ 分类

$$q(X=x)=P(Y=1|X=x)$$

$$\frac{q}{1-q}\in(0,+\infty)\quad\log\frac{q}{1-q}\in(-\infty,+\infty)$$

$$\log\frac{q}{1-q}=ax_1+bx_2+cx_3+d=\beta x$$

$$q=\frac{e^{\beta x}}{1+e^{\beta x}}$$

$$P(Y_1=y_1,\ldots,Y_n=y_n|X_1=x_1,\ldots,X_n=x_n)$$

$$
\begin{aligned}
&= \prod_{i:y_i=1} q(x_i, \beta) \cdot \prod_{i:y_i=0} (1 - q(x_i, \beta)) \\
&= \sum_{i=1}^{n} y_i \log q(x_i; \beta) + (1 - y_i) \log(1 - q(x_i; \beta)) \\
&= \sum \{y_i \beta^T x_i - \log(1 + e^{\beta^T x_i})\}
\end{aligned}
$$