- H-K 算法

H-K 算法是求解 $\boldsymbol{Xw}=\boldsymbol{b}$，式中 $\boldsymbol{b}=(b_1, b_2, \ldots, b_N)^T$，$\boldsymbol{b}$ 的所有分量都是正值。这里要同时计算 $\boldsymbol{w}$ 和 $\boldsymbol{b}$，我们已知 $\boldsymbol{X}$ 不是 $N*N$ 的方阵，通常是行多于列的 $N*(n+1)$阶的长方阵，属于超定方程，因此一般情况下，$\boldsymbol{Xw}=\boldsymbol{b}$ 没有唯一确定解，但可求其线性最小二乘解。

设 $\boldsymbol{Xw}=\boldsymbol{b}$ 的线性最小二乘解为 $\boldsymbol{w}^*$，即使$\|\boldsymbol{Xw}^*-\boldsymbol{b}\|$=极小

采用梯度法，定义准则函数：

$$J(\boldsymbol{w},\boldsymbol{x},\boldsymbol{b})=\frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{w}^T\boldsymbol{x}^i-b_i\right)^2=\frac{1}{2}\sum_{i=1}^{N}\left((\boldsymbol{x}^i)^T\boldsymbol{w}-b_i\right)^2$$
$$=\frac{1}{2}\|\boldsymbol{Xw}-\boldsymbol{b}\|^2=\frac{1}{2}(\boldsymbol{Xw}-\boldsymbol{b})^T(\boldsymbol{Xw}-\boldsymbol{b})$$

当 $\boldsymbol{Xw}=\boldsymbol{b}$ 的条件满足时，J 达到最小值。由于上式中包括的 $\sum_{i=1}^{N}\left(\boldsymbol{w}^T\boldsymbol{x}_i-\boldsymbol{b}_i\right)^2$ 项为两个数量方差的和，且我们将使其最小化，因此也称之为**最小均方误差算法**。

使函数 J 同时对变量 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 求最小。对于 $\boldsymbol{w}$ 的梯度为：

$$\frac{\partial J}{\partial \boldsymbol{w}}=\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{Xw}-\boldsymbol{b})$$

（利用求导公式$\frac{\partial \boldsymbol{Ax}}{\partial \boldsymbol{x}}=\boldsymbol{A}^T$，$\frac{\partial \boldsymbol{x}^T\boldsymbol{Ax}}{\partial \boldsymbol{x}}=\boldsymbol{Ax}+\boldsymbol{A}^T\boldsymbol{x}$）

使$\frac{\partial J}{\partial \boldsymbol{w}}=0$，得 $\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{Xw}-\boldsymbol{b})=0$，从而 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Xw}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{b}$。因为 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 为$(n+1)*(n+1)$阶方阵，因此可求得解：

$$\boldsymbol{w}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{X}^{\#}\boldsymbol{b}$$

这里 $\boldsymbol{X}^{\#}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ 称为 $\boldsymbol{X}$ 的伪逆，$\boldsymbol{X}$ 是 $N*(n+1)$阶的长方阵。

由上式可知，只要求出 $\boldsymbol{b}$ 即可求得 $\boldsymbol{w}$。利用梯度法可求得 $\boldsymbol{b}$ 的迭代公式为：

$$\boldsymbol{b}(k+1)=\boldsymbol{b}(k)-\mathrm{C}\cdot\left(\frac{\partial \mathrm{J}}{\partial \boldsymbol{b}}\right)_{\boldsymbol{b}=\boldsymbol{b}(k)}$$

根据上述约束条件，在每次迭代中，$\boldsymbol{b}(k)$的全部分量只能是正值。由 J 的准则函数式，J 也是正值，因此，当取校正增量 C 为正值时，为保证每次迭代中的 $\boldsymbol{b}(k)$都是正值，应使$\left(\frac{\partial J}{\partial \boldsymbol{b}}\right)_{\boldsymbol{b}=\boldsymbol{b}(k)}$为非正值。在此条件下，准则函数 J 的微分为：

$$-2\left(\frac{\partial J}{\partial \boldsymbol{b}}\right)_{\boldsymbol{b}=\boldsymbol{b}(k)}=(\boldsymbol{Xw}-\boldsymbol{b})+|\boldsymbol{Xw}-\boldsymbol{b}|$$

该式满足以下条件：

若$[\boldsymbol{Xw}(k)-\boldsymbol{b}(k)]>0$，则$-\left(\frac{\partial J}{\partial \boldsymbol{b}}\right)_{\boldsymbol{b}=\boldsymbol{b}(k)}=\boldsymbol{Xw}(k)-\boldsymbol{b}(k)$.

若$[\boldsymbol{Xw}(k)-\boldsymbol{b}(k)]<0$，则$-\left(\frac{\partial J}{\partial \boldsymbol{b}}\right)_{\boldsymbol{b}=\boldsymbol{b}(k)}=0$

由 $\boldsymbol{b}$ 的迭代式和微分，有：

$$\boldsymbol{b}(k+1)=\boldsymbol{b}(k)+\delta\,\boldsymbol{b}(k)$$

$$\delta\,\boldsymbol{b}(k)=C[\boldsymbol{Xw}(k)-\boldsymbol{b}(k)+|\,\boldsymbol{Xw}(k)-\boldsymbol{b}(k)|]$$

将此式代入 $\boldsymbol{w}=\boldsymbol{X}^{\#}\boldsymbol{b}$，有：

$$\boldsymbol{w}(k+1)=\boldsymbol{X}^{\#}\boldsymbol{b}(k+1)=\boldsymbol{X}^{\#}[\boldsymbol{b}(k)+\delta\,\boldsymbol{b}(k)]=\boldsymbol{w}(k)+\boldsymbol{X}^{\#}\,\delta\,\boldsymbol{b}(k)$$

为简化起见，令 $\boldsymbol{e}(k)=\boldsymbol{Xw}(k)-\boldsymbol{b}(k)$，可得 H-K 算法的迭代式。

设初值为 $\boldsymbol{b}(1)$，其每一分量均为正值，则：

$$\boldsymbol{w}(1)=\boldsymbol{X}^{\#}\boldsymbol{b}(1)$$

$$\boldsymbol{e}(k)=\boldsymbol{Xw}(k)-\boldsymbol{b}(k)$$

$$\begin{aligned}\boldsymbol{w}(k+1)&=\boldsymbol{w}(k)+\boldsymbol{X}^{\#}\{\mathrm{C}[\boldsymbol{Xw}(k)-\boldsymbol{b}(k)+|\boldsymbol{Xw}(k)-b(k)|]\}\\&=\boldsymbol{w}(k)+C\boldsymbol{X}^{\#}[\boldsymbol{e}(k)+|\boldsymbol{e}(k)|]\end{aligned}$$

由于

$$\boldsymbol{X}^{\#}\boldsymbol{e}(k) = \boldsymbol{X}^{\#}[\,\boldsymbol{X}\boldsymbol{w}(k) - \boldsymbol{b}(k)] = (\boldsymbol{X}^{T}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}[\boldsymbol{X}\boldsymbol{w}(k) - \boldsymbol{b}(k)]$$

$$= \boldsymbol{w}(k) - \boldsymbol{X}^{\#}\boldsymbol{b}(k) = 0$$

因此

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + C\boldsymbol{X}^{\#}|\boldsymbol{e}(k)|$$

$$\boldsymbol{b}(k+1) = \boldsymbol{b}(k) + C[\boldsymbol{X}\boldsymbol{w}(k) - \boldsymbol{b}(k) + |\boldsymbol{X}\boldsymbol{w}(k) - \boldsymbol{b}(k)|]$$

$$= \boldsymbol{b}(k) + \boldsymbol{C}[\boldsymbol{e}(k) + |\,\boldsymbol{e}(k)|]$$

+

注：常用求导公式：

y是标量，$\boldsymbol{x}=\begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}^T$ 是d维列向量,则$\dfrac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2} & \cdots & \dfrac{\partial y}{\partial x_d} \end{bmatrix}^T$

常用求导公式：

$$\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^T$$

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{x}$$

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{x}$$