PSI-3260 Aplicações de Álgebra Linear

Experiência 8

Método dos Mínimos Quadrados 1 – Parte teórica

O método dos mínimos quadrados (LS - *Least Squares*) é usualmente aplicado para solucionar problemas de otimização por minimização (energia) ou maximização (entropia) de uma determinada função. Originalmente, foi publicado em 1809 por Carl Friedrich Gauss (1777-1855).

1 Aproximação linear dos mínimos quadrados

O ajuste de dados a curvas é um tipo de problema muito frequente em ciências experimentais. Por exemplo, em certos problemas existe necessidade de fazer uma previsão de um dado valor com base em um conjunto de informações previamente observadas. A solução para esse tipo de problema pode ser obtida a partir da resolução de um sistema de equações do tipo

$$\mathbf{A}\mathbf{x} = \mathbf{b}.$$

Especificamente, considere um conjunto de pontos determinados experimentalmente, por exemplo,

$$\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}.$$

Deseja-se obter uma relação matemática $y = f(x)$ entre as variáveis x e y através do *ajuste* de uma curva aos pontos no plano que correspondem aos vários valores de x e de y. Muitas vezes, baseado-se simplesmente no padrão apresentado pelos pontos, escolhemos a forma geral da curva a ser ajustada. Por exemplo, podemos escolher:

1. uma reta: $y = ax + b$,
2. um polinômio quadrático: $y = a + b\,x + c\,x^2$,
3. um polinômio cúbico: $y = a + b\,x + c\,x^2 + d\,x^3$,
4. uma combinação de senóides, exponenciais, entre outras funções.

Como os pontos são obtidos experimentalmente, uma aproximação exata não é possível. A ideia então, é escolher a curva que melhor se ajusta aos dados. Uma primeira tentativa seria considerar uma aproximação linear para a relação dos dados observados. Neste caso, queremos ajustar a reta

$$y = ax + b$$

aos pontos $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ determinados experimentalmente. Apenas quando esses pontos são colineares, a reta passa por todos os n pontos e os coeficientes desconhecidos a e b satisfazem

$$\begin{aligned} y_1 &= a\,x_1 + b \\ y_2 &= a\,x_2 + b \\ &\vdots \\ y_n &= a\,x_n + b. \end{aligned} \tag{1}$$

Por conveniência, vamos escrever o Sistema de Equações (1) na forma matricial

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} b \\ a \end{bmatrix}}_{\mathbf{v}}. \tag{2}$$

Note que neste caso $\mathbf{y} - \mathbf{M}\mathbf{v} = \mathbf{0}$. Se os pontos determinados experimentalmente não são colineares, o sistema (2) é inconsistente, ou seja, é impossível encontrar coeficientes a e b que satisfaçam (2) exatamente e $\mathbf{y} - \mathbf{M}\mathbf{v} \neq \mathbf{0}$. Visando encontrar a melhor solução para o caso inconsistente, vamos representar a diferença entre os dois vetores como

$$\mathbf{e} = \mathbf{y} - \mathbf{M}\mathbf{v}. \tag{3}$$

Os elementos do vetor $\mathbf{e}$ são $e_i = y_i - b - x_i a$ para $i = 1, \cdots, n$. Note que esses escalares podem ser interpretados como a distância vertical entre uma reta $y = ax + b$ os os pontos obtidos experimentalmente (x_i, y_i). Essa distância é uma medida do "desvio" ou "erro" que resulta do ajuste inexato do ponto (x_i, y_i) à reta $y = ax + b$. Para exemplificar, na Figura 1 estão representados um conjunto de pontos $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ e seus correspondentes desvios em relação a uma data reta. A melhor reta para representar o conjunto de pontos não colineares pode ser obtida com a solução dos mínimos quadrados descrita a seguir.

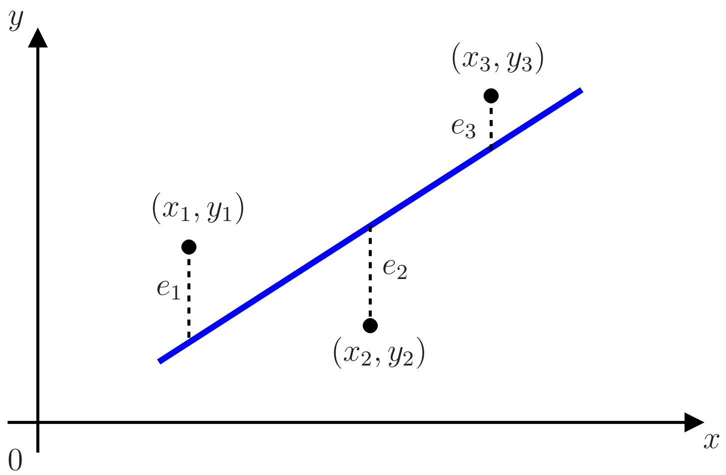

Figura 1: Distância de um conjunto de pontos a uma determinada reta.

Considere a seguinte função

$$\|\mathbf{e}\|^2 = \sum_{i=1}^{n} e_i^2 = \|\mathbf{y} - \mathbf{M}\mathbf{v}\|^2 = \sum_{i=1}^{n} (y_i - a x_i - b)^2. \tag{4}$$

Note que $\|\mathbf{e}\|$ representa a norma Euclidiana do vetor de erros[1]. O método dos mínimos quadrados tem como objetivo obter os coeficientes a e b de modo a assegurar o menor erro quadrático médio, ou seja, tornar o vetor $\mathbf{e}$ de menor norma possível. Assim, derivando a Equação (4) em relação a cada um dos elementos do vetor de coeficientes $\mathbf{v}$ obtemos

$$\begin{aligned} \frac{\partial \sum_{i=1}^{n} e_i^2}{\partial b} &= 2\sum_{i=1}^{n} e_i \frac{\partial e_i}{\partial b} = -2\sum_{i=1}^{n} e_i \\ \frac{\partial \sum_{i=1}^{n} e_i^2}{\partial a} &= 2\sum_{i=1}^{n} e_i \frac{\partial e_i}{\partial a} = -2\sum_{i=1}^{n} e_i x_i. \end{aligned} \tag{5}$$

De forma compacta podemos escrever

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \mathbf{v}} = -2\sum_{i=1}^{n} \begin{bmatrix} 1 \\ x_i \end{bmatrix} e_i = -2\mathbf{M}^T\mathbf{e} = -2\mathbf{M}^T(\mathbf{y} - \mathbf{M}\mathbf{v}), \tag{6}$$

em que $(\cdot)^T$ representa a operação de transposição da matriz $\mathbf{M}$. Igualando (6) ao vetor nulo

$$-2\mathbf{M}^T(\mathbf{y} - \mathbf{M}\mathbf{v}^*) = \mathbf{0}, \tag{7}$$

obtemos

$$\mathbf{M}^T\mathbf{M}\mathbf{v}^* = \mathbf{M}^T\mathbf{y}. \tag{8}$$

Portanto, os coeficientes a e b que satisfazem (8), denotados como a^* e b^*, representam a solução dos mínimos quadrados, ou seja, minimizam a norma do vetor do erro.

Se $\mathbf{M}^T\mathbf{M}$ for invertível, o vetor

$$\mathbf{v}^* = \begin{bmatrix} b^* \\ a^* \end{bmatrix}$$

pode ser calculado diretamente como

$$\boxed{\mathbf{v}^* = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{y}.} \tag{9}$$

Essa equação expressa a unicidade da solução $\mathbf{v} = \mathbf{v}^*$ na igualdade (8). Note que o vetor $\mathbf{v}^*$ é aquele que fornece a menor norma do vetor $\|\mathbf{e}\|$. Assim, se os pontos experimentais são não colineares, então, existe um único ajuste linear dos mínimos quadrados a estes pontos, dados por

$$y = b^* + a^* x.$$

[1]Lembre que a definição da norma Euclidiana e de outras normas já foram tratadas na aplicação da aula de Medidas de Distância e Volume.

Exercício 1

Seja o conjunto de pontos obtidos experimentalmente

$$\{(0,1),\ (1,3),\ (2,4)\}.$$

Determine a melhor relação linear segundo o critério dos mínimos quadrados para o ajuste desse conjunto de pontos.

2 Ajuste polinomial dos mínimos quadrados

A técnica descrita de ajuste linear dos mínimos quadrados pode ser facilmente generalizada para ajustar um polinômio de grau m qualquer a um conjunto de pontos observados. Especificamente, considere o polinômio de grau m

$$y = a_0 + a_1 x + \cdots + a_m x^m$$

e o conjunto de n pontos experimentais

$$\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}.$$

Substituindo os n pares de pontos observados na equação polinomial temos

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}}_{\mathbf{v}} \tag{10}$$

Como no caso do ajuste linear dos mínimos quadrados, a minimização da norma do erro quadrático

$$\|\mathbf{e}\|^2 = \|\mathbf{y} - \mathbf{M}\mathbf{v}\|^2$$

ocorre quando o vetor de coeficientes satisfaz a igualdade (9). Entretanto, a matriz $\mathbf{M}$ e os vetores $\mathbf{y}$ e $\mathbf{v}$ são definidos genericamente como em (10).

Exercício 2

Encontre a parábola segundo o método dos mínimos quadrados que melhor se ajusta aos seguintes pontos obtidos experimentalmente

$$\{(-1,14),\ (0,-5),\ (1,-4),\ (2,1),\ (3,22)\}.$$

Você pode usar o Matlab para resolver esse exercício.

3 Interpretação geométrica

A definição do vetor de erros $\mathbf{e} = \mathbf{y} - \mathbf{Mv}$ sugere que $\mathbf{Mv}$ pode ser interpretado como uma estimativa do vetor $\mathbf{y}$, assim definimos

$$\hat{\mathbf{y}} = \mathbf{Mv}.$$

O ponto de mínimo da função (4) ocorre quando $\mathbf{M}^T\mathbf{e}^* = \mathbf{0}$ em que $\mathbf{e}^* = \mathbf{y} - \hat{\mathbf{y}}^*$. Em outras palavras, o vetor de erros que obedece a condição dos mínimos quadrados é ortogonal as colunas da matriz $\mathbf{M}$.

O vetor de coeficientes que satisfaz a condição dos mínimos quadrados obedece a relação (9). Portanto, a melhor estimativa de $\mathbf{y}$ segundo o critério dos mínimos quadrados é

$$\hat{\mathbf{y}}^* = \mathbf{Mv}^* = \underbrace{\mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T}_{\boldsymbol{\mathcal{P}}}\mathbf{y}$$

A matriz

$$\boldsymbol{\mathcal{P}} = \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$$

projeta o vetor de resposta desejada no espaço formado pelas colunas da matriz de dados, sendo chamada de operador de projeção. Cabe aqui notar que tal operador tem dimensão $n \times n$, é idempotente[2] e simétrico. Neste caso, como $\hat{\mathbf{y}}^* = \boldsymbol{\mathcal{P}}\mathbf{y}$, o vetor de erro pode ser expresso como

$$\mathbf{e}^* = \mathbf{y} - \hat{\mathbf{y}}^* = \mathbf{y} - \boldsymbol{\mathcal{P}}\mathbf{y} = \boldsymbol{\mathcal{P}}^{\perp}\mathbf{y}$$

sendo $\boldsymbol{\mathcal{P}}^{\perp} = \mathbf{I} - \boldsymbol{\mathcal{P}}$ o complementar do operador de projeção, que projeta o vetor de medidas no espaço ortogonal às colunas da matriz de dados.

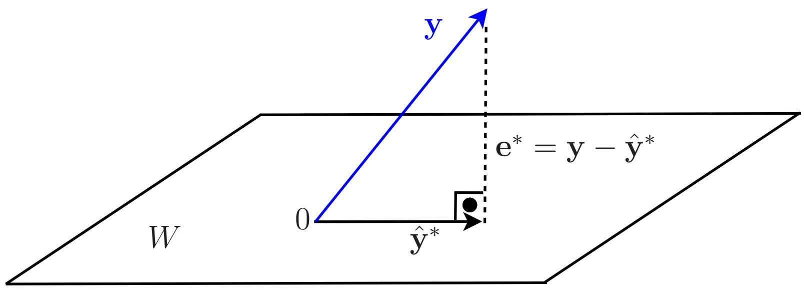

Figura 2: Representação do princípio da ortogonalidade: $\hat{\mathbf{y}}^*$ é a projeção do vetor $\mathbf{y}$ no espaço W formado pelas colunas da matriz $\mathbf{M}$ e $\mathbf{e}^* = \mathbf{y} - \hat{\mathbf{y}}^*$ é um vetor ortogonal a esse espaço.

O princípio da ortogonalidade é resultado da relação $\mathbf{M}^T\mathbf{e}^* = \mathbf{0}$, ou seja, o vetor de erros que obedece à condição dos mínimos quadrados é ortogonal às colunas da matriz $\mathbf{M}$. Uma consequência do princípio da ortogonalidade é $\mathbf{e}^* \perp \hat{\mathbf{y}}^*$, ou seja, o vetor de erros com a menor norma é ortogonal à projeção do vetor $\mathbf{y}$ no espaço formado pelas colunas da matriz $\mathbf{M}$. Na

[2]Uma matriz A é idempotente quando $AA = A$.

Figura 2, ilustramos o efeito do princípio da ortogonalidade e seu corolário. Cabe notar que o operador $\boldsymbol{\mathcal{P}}$ faz uma projeção de $\mathbf{y}$ no espaço formado pelas colunas da matriz de dados resultando no vetor $\hat{\mathbf{y}}$ ótimo e o operador $\boldsymbol{\mathcal{P}}^{\perp}$ faz uma projeção de $\mathbf{y}$ em um espaço que é ortogonal àquele formado pelas colunas da matriz de dados resultando no vetor $\mathbf{e}$ que tem norma mínima.