Matemática Computacional
MEBiol, MEBiom e MEFT
Aula 15 - Ajustamento de dados discretos e aproximação de funções

Ana Leonor Silvestre

Instituto Superior Técnico, $1^{\underline{o}}$ *Semestre, 2020/2021*

Sumário da Aula 15

Cap. 4 - Ajustamento de dados discretos e aproximação de funções

Método dos mínimos quadrados.

Algumas considerações sobre interpolação polinomial e ajustamento de dados através do método dos mínimos quadrados

- O problema com a interpolação polinomial de grau elevado é a ocorrência de grandes oscilações entre nós de interpolação. Podemos dizer que há demasiado esforço para tentar ajustar todos os pontos dados e acaba por se perder a estrutura do conjunto de dados. Como consequência, os resultados para as previsões podem ser bastante imprecisos.
  A este fenómeno chama-se overfitting: estamos a tentar ajustar um polinómio de grau maior do que o necessário para fazer o ajustamento.

Exemplo - Comparação entre interpolação polinomial e regressão linear

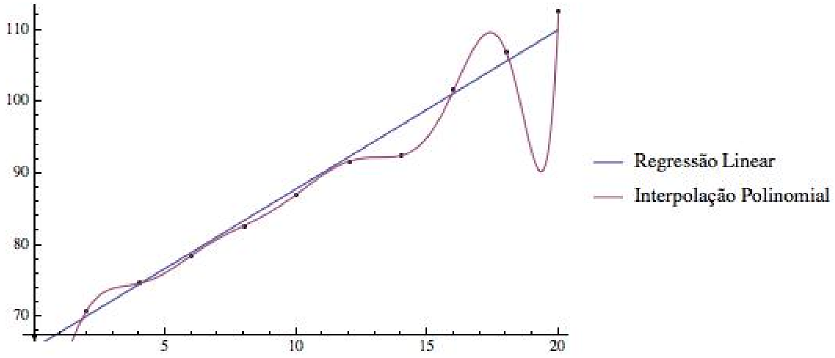

Figura: Funções de ajustamento

Algumas considerações sobre interpolação polinomial e ajustamento de dados através do método dos mínimos quadrados

- Interpolação, ou, mais frequentemente, ajustamento de dados discretos, é uma ferramenta de aprendizagem automática (machine learning). É mais simples do que a maioria dos algoritmos de aprendizagem automática, mas é útil em variadas situações.
- A regressão linear é um algoritmo de aprendizagem automática baseado na aprendizagem supervisionada. O objectivo da aprendizagem supervisionada é *aprender* a prever uma ou várias variáveis como função de outras variáveis, a partir de um conjunto de exemplos.

Ajustamento de dados discretos através do método dos mínimos quadrados

Pretende-se ajustar uma função g a um conjunto de pares de números reais

x_k	x_1	x_2	$\cdots$	x_n
y_k	y_1	y_2	$\cdots$	y_n

Podemos considerar que os dados correspondem ao registo de uma função desconhecida $y : \{x_{1,\dots,}x_n\} \to \mathbb{R}$, tal que

$$y(x_k) = y_k, \quad k = 1, \dots, n.$$

Segundo o critério dos mínimos quadrados, uma vez decidida a forma da função aproximante g, a qual dependerá de um número finito de parâmetros reais, determina-se os valores desses parâmetros de modo que a soma dos desvios quadrados

$$D := \sum_{k=1}^{n} [y_k - g(x_k)]^2 = \sum_{k=1}^{n} [y(x_k) - g(x_k)]^2$$

seja mínima.

Exemplos

- O ajustamento de uma reta

$$g(x) = c_1 + c_2 x$$

a um conjunto de pontos segundo o critério dos mínimos quadrados é designado por *regressão linear*.

- No caso de dados periódicos ou quase periódicos, como na observação de marés ou mudanças climáticas, são usados polinómios trigonométricos

$$g(x) = \sum_{i=1}^{m} [a_i \cos(\alpha_i x) + b_i \sin(\beta_i x)]$$

para efetuar o ajustamento.

Método dos mínimos quadrados

No espaço das funções reais definidas no conjunto $\{x_{1,\ldots,}x_n\}$, consideramos o produto interno

$$\langle \varphi, \psi \rangle := \sum_{k=1}^{n} \varphi(x_k)\psi(x_k)$$

Vamos estudar o caso mais simples, em que a função de ajustamento é combinação linear de funções contínuas $\phi_1, \ldots, \phi_m$, com $m \leq n$, selecionadas *a priori*:

$$g(x) = \sum_{i=1}^{m} c_i \phi_i(x)$$

As funções $\phi_1, \ldots, \phi_m$ são designadas por *funções de base* para o ajustamento.

Método dos mínimos quadrados

Vejamos então em que condições é que a função

$$D := \sum_{k=1}^{n} [y_k - g(x_k)]^2 = \|y - g\|^2 = \langle y - g, y - g \rangle$$

tem um ponto de mínimo e como o determinar. Tem-se

$$\begin{aligned} D(c_1, ..., c_m) &= \langle y, y \rangle - 2\langle y, g \rangle + \langle g, g \rangle = \\ &= \langle y, y \rangle - 2 \left\langle y, \sum_{i=1}^{m} c_i \phi_i \right\rangle + \left\langle \sum_{i=1}^{m} c_i \phi_i, \sum_{j=1}^{m} c_j \phi_j \right\rangle = \\ &= \langle y, y \rangle - 2 \sum_{i=1}^{m} c_i \langle y, \phi_i \rangle + \sum_{i=1}^{m} \sum_{j=1}^{m} c_i c_j \langle \phi_i, \phi_j \rangle \end{aligned}$$

$D : \mathbb{R}^m \to \mathbb{R}$ é uma função quadrática

A função soma dos desvios quadrados

$$D : \mathbb{R}^m \to \mathbb{R}$$

$$D(c_1, ..., c_m) = \langle y, y \rangle - 2\sum_{i=1}^{m} c_i \langle y, \phi_i \rangle + \sum_{i=1}^{m}\sum_{j=1}^{m} c_i c_j \langle \phi_i, \phi_j \rangle$$

$$\frac{\partial D}{\partial c_i}(c) = -2\langle y, \phi_i \rangle + 2\sum_{j=1}^{m} c_j \langle \phi_i, \phi_j \rangle, \quad i = 1, ..., m$$

$$\frac{\partial^2 D}{\partial c_j \partial c_i}(c) = 2\langle \phi_i, \phi_j \rangle, \quad i, j = 1, ..., m$$

Minimização da soma dos desvios quadrados

Começamos por determinar os pontos de estacionaridade de D. Trata-se dos pontos $c^* = (c_1^*, ..., c_m^*) \in \mathbb{R}^m$ tais que $\nabla D(c^*) = 0$.

$$\nabla D(c) = 0$$

$$\iff$$

$$\frac{\partial D}{\partial c_i}(c) = 0, \quad i = 1, ..., m$$

$$\iff$$

$$-2\langle y, \phi_i \rangle + 2 \sum_{j=1}^{m} c_j \langle \phi_i, \phi_j \rangle = 0, \quad i = 1, ..., m$$

$$\iff$$

$$\sum_{j=1}^{m} c_j \langle \phi_i, \phi_j \rangle = \langle y, \phi_i \rangle, \quad i = 1, ..., m.$$

Sistema normal

Os pontos de estacionaridade de D, se existirem, são solução do sistema linear

$$\sum_{j=1}^{m} c_j \langle \phi_i, \phi_j \rangle = \langle y, \phi_i \rangle, \quad i = 1, ..., m.$$

Em notação matricial, o sistema escreve-se

$$\begin{bmatrix} \langle \phi_1, \phi_1 \rangle & \dots & \langle \phi_1, \phi_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \phi_m, \phi_1 \rangle & \dots & \langle \phi_m, \phi_m \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \langle y, \phi_1 \rangle \\ \vdots \\ \langle y, \phi_m \rangle \end{bmatrix}$$

A este sistema chama-se *sistema normal* ou *sistema de equações normais*. Note-se que, devido à simetria dos produtos internos, $\langle \varphi, \psi \rangle = \langle \psi, \varphi \rangle$, a matriz do sistema normal é simétrica.

Método dos mínimos quadrados

Devemos agora estabelecer condições para que este sistema tenha uma e uma só solução, c^*, e para que um tal c^* seja ponto de mínimo (absoluto) de D.
A matriz Hessiana de D nesse ponto deverá ser definida positiva. Como já vimos,

$$\frac{\partial^2 D}{\partial c_j \partial c_i}(c) = 2\langle \phi_i, \phi_j \rangle, \quad i, j = 1, ..., m$$

pelo que a matriz Hessiana de D é independente de c e dada por

$$H = 2 \begin{bmatrix} \langle \phi_1, \phi_1 \rangle & \dots & \langle \phi_1, \phi_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \phi_m, \phi_1 \rangle & \dots & \langle \phi_m, \phi_m \rangle \end{bmatrix} = 2M$$

onde M designa a matriz do sistema normal.

Matriz do sistema normal

A matriz do sistema normal é da forma $M = B^\top B$, com

$$B = \begin{bmatrix} \phi_1(x_1) & \dots & \phi_m(x_1) \\ \phi_1(x_2) & \dots & \phi_m(x_2) \\ \vdots & \dots & \vdots \\ \phi_1(x_n) & \dots & \phi_m(x_n) \end{bmatrix}.$$

Tem-se também

$$\begin{bmatrix} \langle y, \phi_1 \rangle \\ \vdots \\ \langle y, \phi_m \rangle \end{bmatrix} = B^\top \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Método dos mínimos quadrados - Algoritmo

$$\begin{bmatrix} \langle \phi_1, \phi_1 \rangle & \dots & \langle \phi_1, \phi_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \phi_m, \phi_1 \rangle & \dots & \langle \phi_m, \phi_m \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \langle y, \phi_1 \rangle \\ \vdots \\ \langle y, \phi_m \rangle \end{bmatrix}$$

$$g(x) = \sum_{i=1}^{m} c_i \phi_i(x)$$

onde

$$\langle \phi_i, \phi_j \rangle = \sum_{k=1}^{n} \phi_i(x_k) \phi_j(x_k), \quad i, j = 1, ..., m$$

$$\langle y, \phi_i \rangle = \sum_{k=1}^{n} y_k \phi_i(x_k), \quad i = 1, ..., m$$

Método dos mínimos quadrados - Algoritmo

$$\boxed{B^\top B c = B^\top y}$$

onde

$$B = \begin{bmatrix} \phi_1(x_1) & \dots & \phi_m(x_1) \\ \phi_1(x_2) & \dots & \phi_m(x_2) \\ \vdots & \dots & \vdots \\ \phi_1(x_n) & \dots & \phi_m(x_n) \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Será que estamos de facto a calcular o ponto de mínimo?

Se a matriz

$$M = \begin{bmatrix} \langle \phi_1, \phi_1 \rangle & \dots & \langle \phi_1, \phi_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \phi_m, \phi_1 \rangle & \dots & \langle \phi_m, \phi_m \rangle \end{bmatrix}$$

for definida positiva então

(i) o sistema de equações normais terá uma e uma só solução $c^* = (c_1^*, ..., c_m^*)$, porque uma matriz definida positiva é não singular;

Será que estamos de facto a calcular o ponto de mínimo?

(ii) c^* será ponto de mínimo absoluto de D, pois, pela fórmula de Taylor de segunda ordem para campos escalares,

$$D(c) = D(c^*) + \nabla D(c^*) \cdot (c - c^*) + \frac{1}{2}(c - c^*)^\top H(c - c^*)$$

$$= D(c^*) + (c - c^*)^\top M(c - c^*) > D(c^*)$$

para qualquer $c \in \mathbb{R}^m \setminus \{c^*\}$.

Assim, basta que a matriz M do sistema normal seja definida positiva para que D tenha um e um só ponto de mínimo.

Será que estamos de facto a calcular o ponto de mínimo?

Como $M = B^\top B$, M é certamente semi-definida positiva, e portanto basta que M seja não-singular para garantir que M é definida positiva.

Teorema

A matriz M é definida positiva se e só se as funções ϕ_1,...,ϕ_m são linearmente independentes no conjunto $\{x_1, ..., x_n\}$.