Nhận dạng mẫu (Pattern Recognition)

Hồi quy tuyến tính (Linear regression)

By Hoàng Hữu Việt

Email: viethh@vinhuni.edu.vn

Viện Kỹ thuật và Công nghệ, Đại học Vinh

Vinh, 5/2019

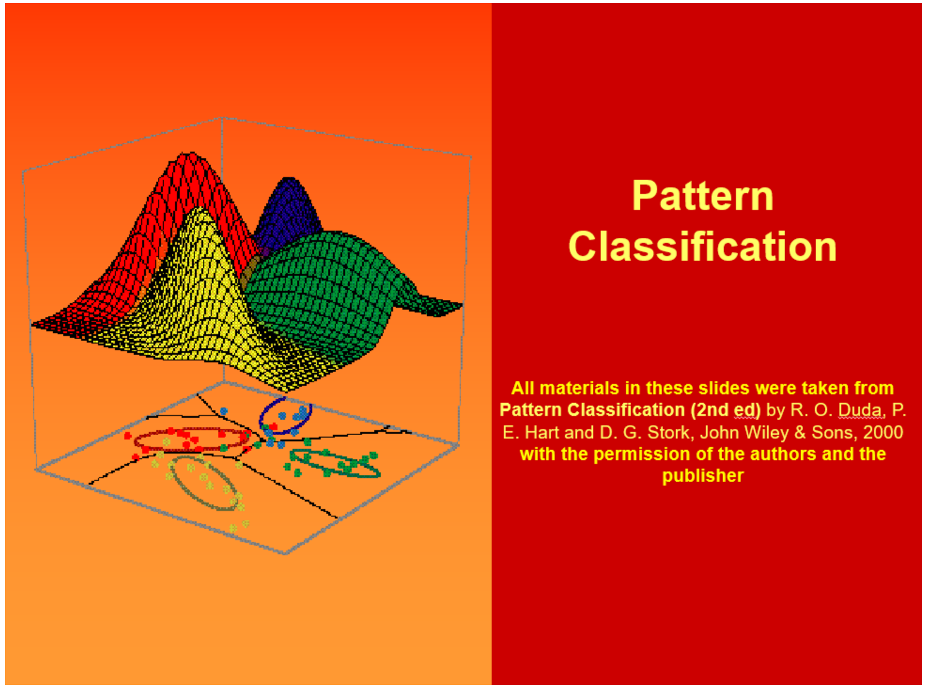
Tài liệu tham khảo

- Tài liệu chính

  [1] Richard O.Duda, Peter E. Hart, David G.Stock. Pattern Classification, 2nd, Wiley, 2001.

  [2]. Milos Hauskrecht. Machine Learning, 2010.

  https://people.cs.pitt.edu/milos/courses/cs2750-Spring2010/



Nội dung

- Hồi quy tuyến tính
- Tối ưu hồi quy tuyến tính
- Phương pháp giả nghịch đảo
- Phương pháp gradient descent
- Mở rộng mô hình tuyến tính

Hồi quy tuyến tính

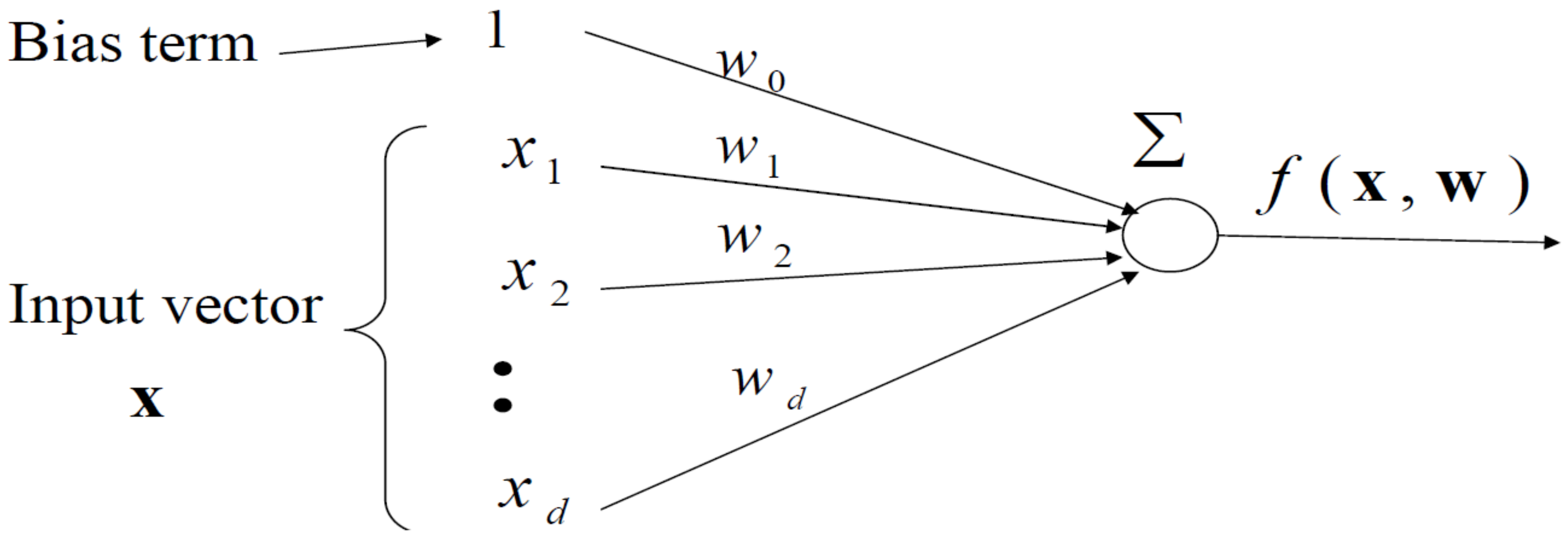
- Học có hướng dẫn:
  - Dữ liệu: $D = \{D_1, D_2, \ldots, D_n\}$ là một tập gồm n mẫu
    - Mỗi mẫu $D_i = (x_i, y_i)$, $i = 1, 2, \ldots, n$.
    - Mẫu vào x_i là một vector kích thước d: $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$.
    - Mẫu ra y_i.
  - Mục tiêu: học một hàm f: $X \rightarrow Y$ sao cho $y_i \approx f(x_i)$, $i = 1, 2, \ldots, n$.
  - Hồi quy (regression): Y liên tục.
  - Phân lớp (classification): Y rời rạc.

Hồi quy tuyến tính

- Hàm $f: X \rightarrow Y$ là một kết hợp tuyến tính của các thành phần dữ liệu vào:

$$f(x) = w_0 + w_1x_1 + w_2x_2 + \ldots + w_dx_d = w_0 + \sum_{j=1}^{d} w_jx_j,$$

trong đó $w_0, w_1, \ldots, w_d$ là các tham số (parameter) hay trọng số (weight).

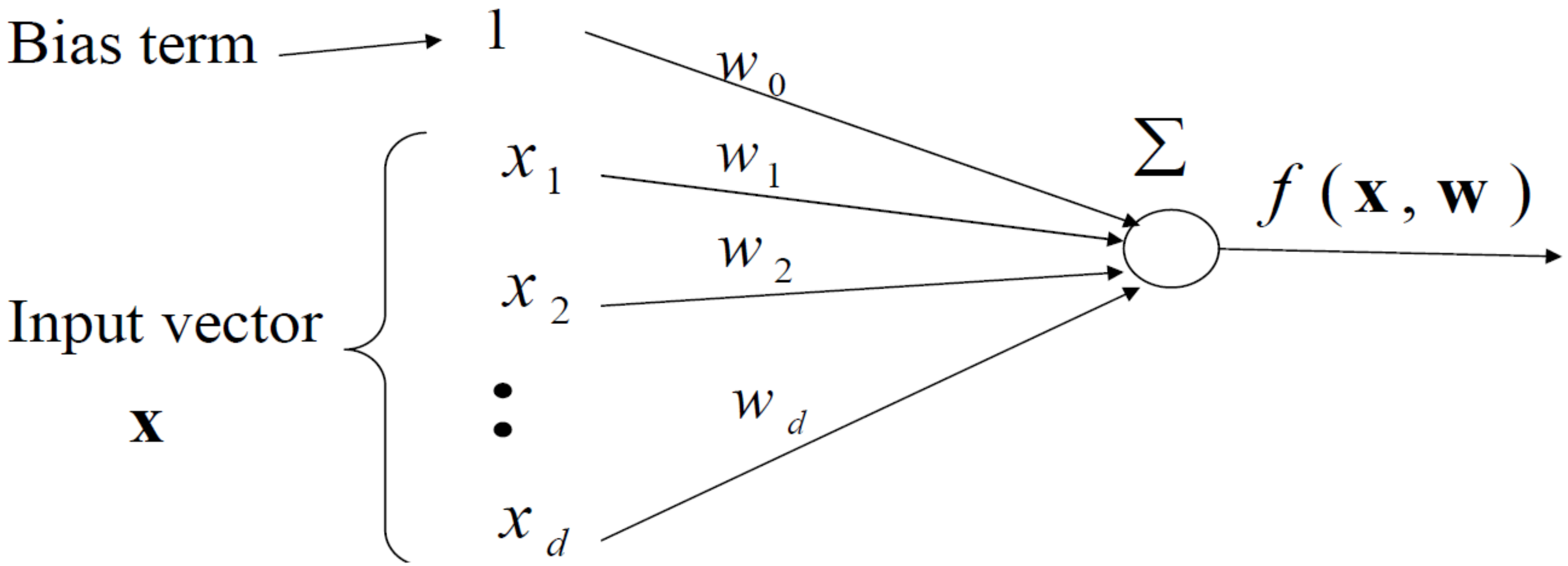


Hồi quy tuyến tính

- Biểu diễn dạng vector:

$$f(x) = w_0 + w_1x_1 + w_2x_2 + \ldots + w_dx_d = w^Tx,$$

trong đó $x = (1, x_1, x_2, \ldots, x_d)^T$ là vector đầu vào và $w = (w_0, w_1, w_2, \ldots, w_d)^T$ là vector tham số (parameter) hay trọng số.



Hàm lỗi

- Dữ liệu $D_i = (x_i, y_i)$, $i = 1, 2, \ldots, n$.
- Mục tiêu là cần tìm một hàm $f(x_i)$ sao cho $y_i \approx f(x_i)$, $i = 1, 2, \ldots, n$.
- Hàm lỗi (error function): sai số của dự đoán so với đầu ra thực tế.
  - Lỗi bình phương trung bình (mean-squared error - MSE):

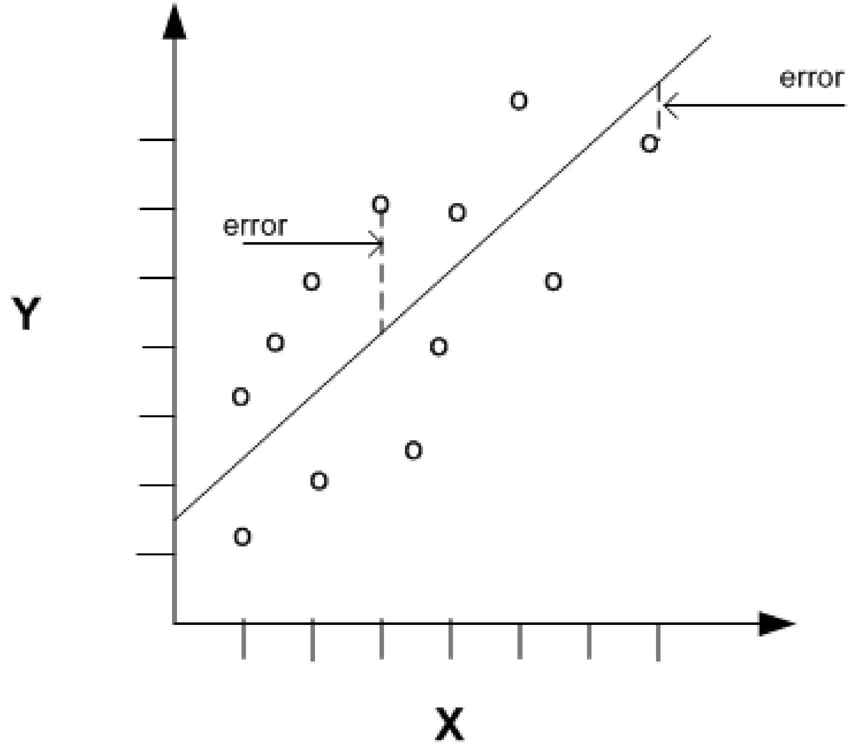
$$J_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

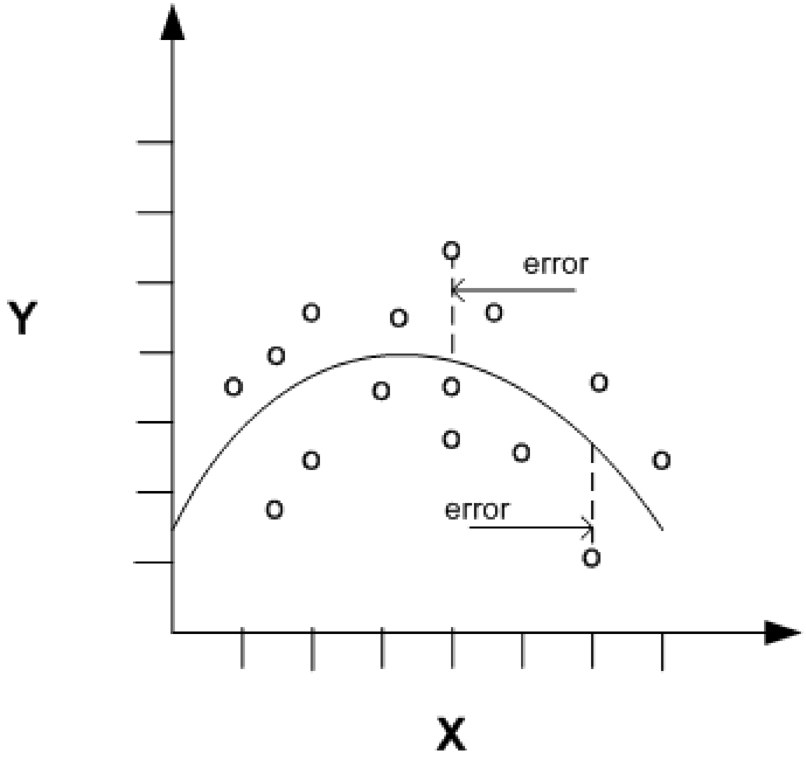
- Học (learning): tìm tập trọng số $w = (w_0, w_1, w_2, \ldots, w_d)^T$ để tối thiểu hóa hàm lỗi J_n.

Hàm lỗi

- Ví dụ dữ liệu vào một chiều $x = (x_1)$

$$J_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

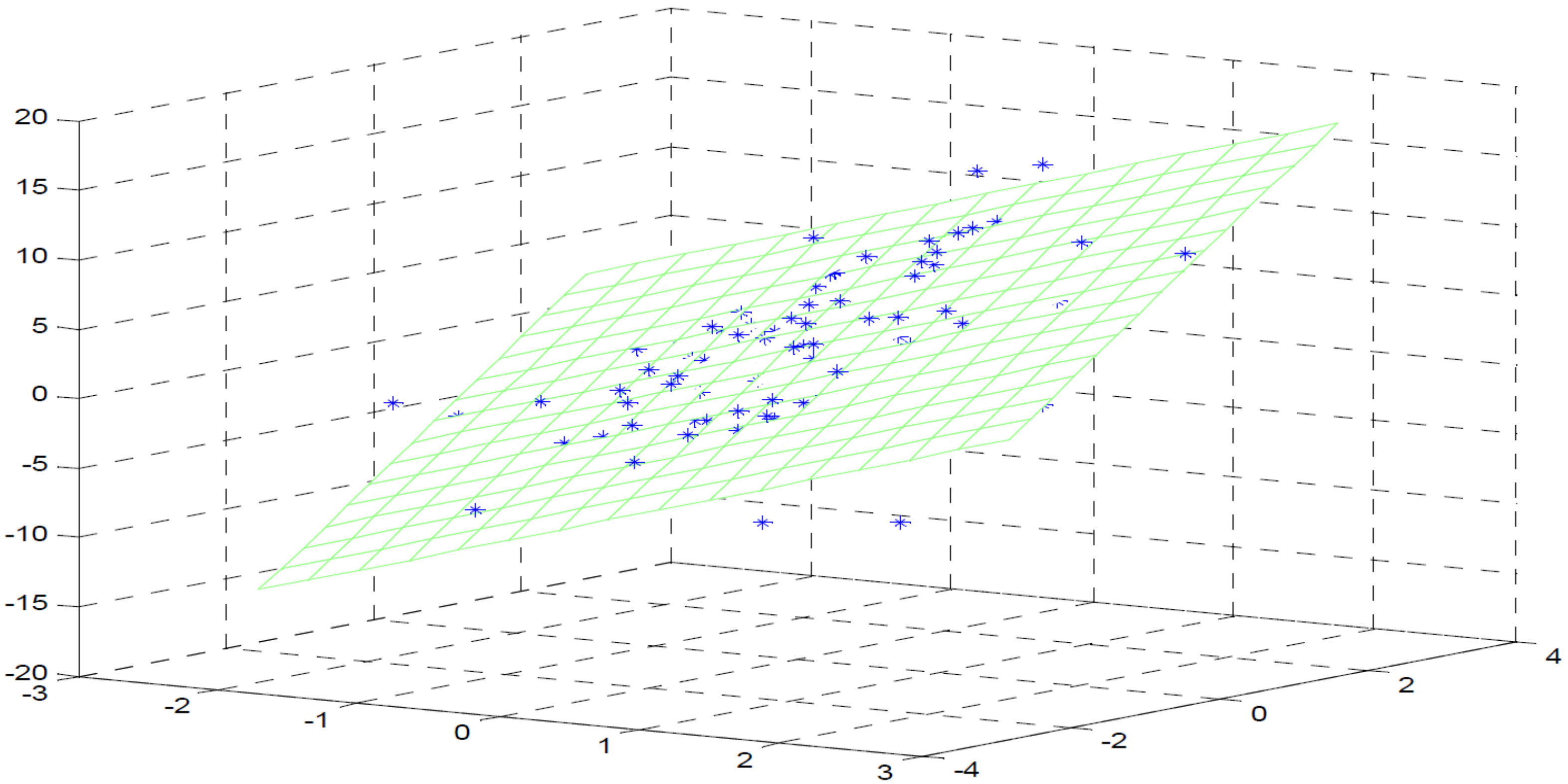




Hàm lỗi

- Dữ liệu vào 2 chiều $x = (x_1, x_2)$

$$J_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$



Tối ưu hồi quy tuyến tính

- Cần tìm vector tham số w để tối thiểu hóa hàm lỗi:

$$J_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^t x_i)^2.$$

- Để tìm vector tham số, đạo hàm của hàm lỗi tương ứng với các tham số phải bằng 0.

$$\frac{\partial}{\partial w_j}J_n(w) = -\frac{2}{n}\sum_{i=1}^{n}(y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id})x_{ij} = 0.$$

Tối ưu hồi quy tuyến tính

- $\nabla_w J_n(w) = 0$ gồm một tập các phương trình:

$$\frac{\partial}{\partial w_0} J_n(w) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id}) x_{i0} = 0$$

$$\frac{\partial}{\partial w_1} J_n(w) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id}) x_{i1} = 0$$

...

$$\frac{\partial}{\partial w_j} J_n(w) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id}) x_{ij} = 0$$

...

$$\frac{\partial}{\partial w_d} J_n(w) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id}) x_{id} = 0$$

Tối ưu hồi quy tuyến tính

- $\nabla_w J_n(w) = 0$ gồm một tập các phương trình:

$$\sum_{i=1}^{n}(y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id})x_{i0} = 0$$

$$\sum_{i=1}^{n}(y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id})x_{i1} = 0$$

$$\ldots$$

$$\sum_{i=1}^{n}(y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id})x_{ij} = 0$$

$$\ldots$$

$$\sum_{i=1}^{n}(y_i - w_0 x_{i0} - w_1 x_{i1} - \cdots - w_d x_{id})x_{id} = 0$$

Tối ưu hồi quy tuyến tính

- $\nabla_w J_n(w) = 0$ gồm một tập các phương trình:

$$
\begin{aligned}
&w_0 \sum_{i=1}^{n} x_{i0}x_{i0} + w_1 \sum_{i=1}^{n} x_{i1}x_{i0} + \cdots + w_d \sum_{i=1}^{n} x_{id}x_{i0} = \sum_{i=1}^{n} y_i x_{i0} \\
&w_0 \sum_{i=1}^{n} x_{i0}x_{i1} + w_1 \sum_{i=1}^{n} x_{i1}x_{i1} + \cdots + w_d \sum_{i=1}^{n} x_{id}x_{i1} = \sum_{i=1}^{n} y_i x_{i1} \\
&\ldots \\
&w_0 \sum_{i=1}^{n} x_{i0}x_{ij} + w_1 \sum_{i=1}^{n} x_{i1}x_{ij} + \cdots + w_d \sum_{i=1}^{n} x_{id}x_{ij} = \sum_{i=1}^{n} y_i x_{ij} \\
&\ldots \\
&w_0 \sum_{i=1}^{n} x_{i0}x_{id} + w_1 \sum_{i=1}^{n} x_{i1}x_{id} + \cdots + w_d \sum_{i=1}^{n} x_{id}x_{id} = \sum_{i=1}^{n} y_i x_{id}
\end{aligned}
$$

- Đây là hệ phương trình có d+1 ẩn $w_0, w_1, \ldots, w_1$: $Aw = b$.

Tối ưu hồi quy tuyến tính

- Tập trọng số tối ưu thõa mãn:

$$\nabla_w(J_n(w)) = -\frac{2}{n}\sum_{i=1}^{n}(y_i - w^T x_i)x_i = \bar{0}$$

dẫn tới một hệ phương trình tuyến tính với $d+1$ ẩn:

$$\underbrace{w_0\sum_{i=1}^{n}x_{i,0}x_{i,j} + w_1\sum_{i=1}^{n}x_{i,1}x_{i,j} + \ldots + w_j\sum_{i=1}^{n}x_{i,j}x_{i,j} + \ldots + w_d\sum_{i=1}^{n}x_{i,d}x_{i,j} = \sum_{i=1}^{n}y_i x_{i,j}}_{\mathbf{A}\mathbf{w}=\mathbf{b}}$$

- Nghiệm: $w = A^{-1}b$, như vậy cần tìm ma trận nghịch đảo!

Tối ưu hồi quy tuyến tính

- Ví dụ cho $D = \{(1,2), (2,1),(2,3),(3,2),(4,3),(5,2)\}$. Tìm đường thẳng để tối thiểu hóa khoảng cách từ các điểm đến đường thẳng đó.
- Ký hiệu mỗi mẫu $D_i = (x_{i1}, y_1)$ (số chiều $d = 1$).
- Ký hiệu đường thẳng: $f(x) = w_0x_0 + w_1x_1$ với $x_0 = 1$.
- Định nghĩa hàm lỗi:

$$J_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 = \frac{1}{6}\sum_{i=1}^{6}(y_i - f(x_i))^2 = \frac{1}{6}\sum_{i=1}^{6}(y_i - w_0x_{i0} - w_1x_{i1})^2.$$

- Tính đạo hàm theo w_0 và w_1:

$$\frac{\partial}{\partial w_0}J_n(w) = -\frac{2}{6}\sum_{i=1}^{6}(y_i - w_0x_{i0} - w_1x_{i1})$$

$$\frac{\partial}{\partial w_1}J_n(w) = -\frac{2}{6}\sum_{i=1}^{6}(y_i - w_0x_{i0} - w_1x_{i1})x_{i1}$$

Tối ưu hồi quy tuyến tính

- Giải $\frac{\partial}{\partial w_0} J_n(w) = 0$ và $\frac{\partial}{\partial w_1} J_n(w) = 0$:

  $-\frac{2}{6}\sum_{i=1}^{6}(y_i - w_0 x_{i0} - w_1 x_{i1}) = 0$ và $-\frac{2}{6}\sum_{i=1}^{6}(y_i - w_0 x_{i0} - w_1 x_{i1})x_{i1} = 0$

  - Tương đương với:

  $\sum_{i=1}^{6}(y_i - w_0 x_{i0} - w_1 x_{i1}) = 0$ và $\sum_{i=1}^{6}(y_i - w_0 x_{i0} - w_1 x_{i1})x_{i1} = 0$

  - Tương đương với:

  $w_0 \sum_{i=1}^{6} x_{i0} + w_1 \sum_{i=1}^{6} x_{i1} = \sum_{i=1}^{6} y_i$ và $w_0 \sum_{i=1}^{6} x_{i0}x_{i1} + w_1 \sum_{i=1}^{6} x_{i1}x_{i1} = \sum_{i=1}^{6} y_i x_{i1}$

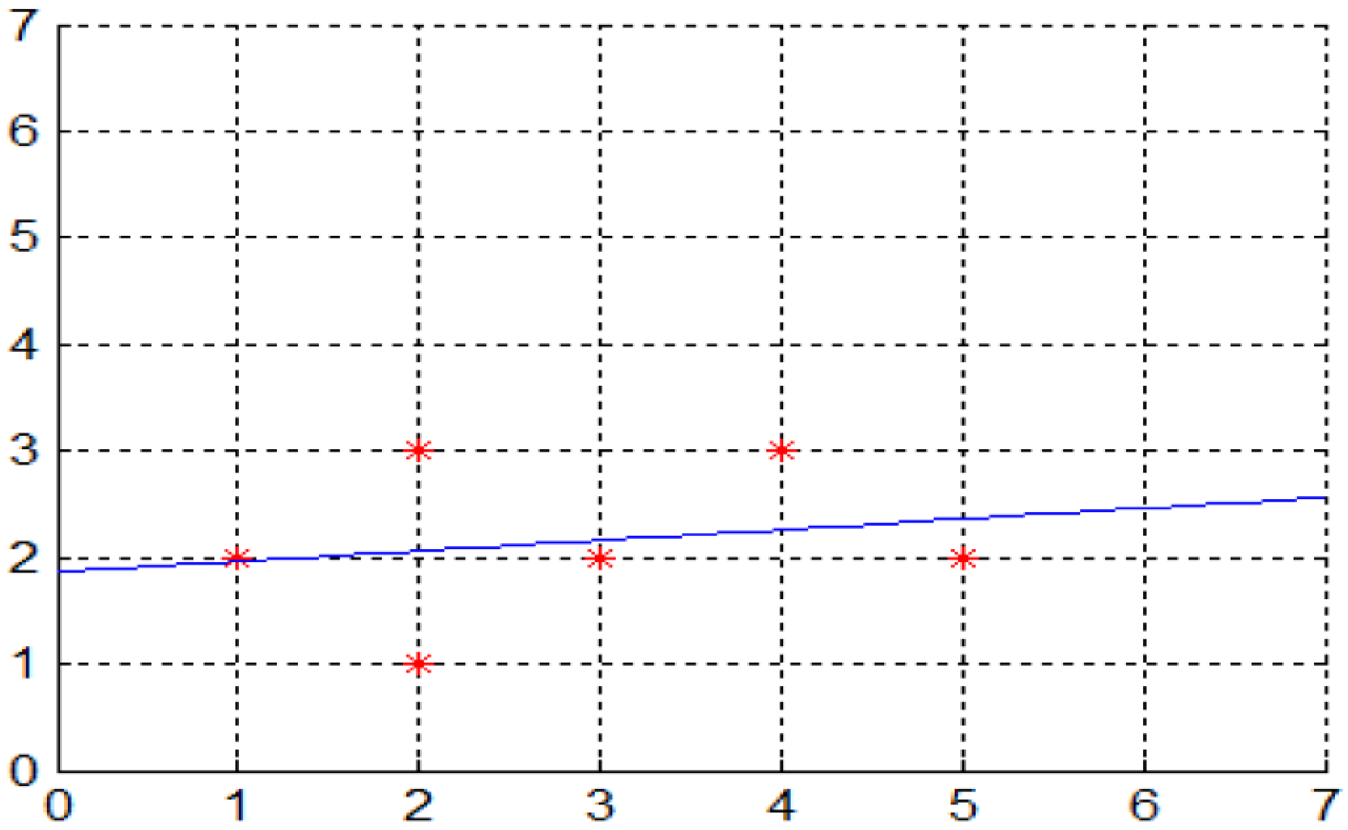
  - Thay $\mathcal{D} = \{(1,2),(2,1),(2,3),(3,2),(4,3),(5,2)\}$ và $x_{i0} = 1$:

  $6w_0 + 17w_1 = 13$ và $17w_0 + 59w_1 = 38 \Rightarrow w_0 = 1.86, w_1 = 0.10$

  - Phương trình đường thẳng: $f(x) = 1.8615 + 0.1077x_1$.

Tối ưu hồi quy tuyến tính

- Ví dụ cho $D = \{(1,2), (2,1),(2,3),(3,2),(4,3),(5,2)\}$. Phương trình đường thẳng $f(x) = 1.8615 + 0.1077x_1$.



Giải bằng giả nghịch đảo

- Hàm cần học:

$$f(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d = w^t x = x^t w,$$

trong đó $x = (1, x_1, \ldots, x_d)^t$ và $w = (w_0, w_1, \ldots, w_d)^t$.

- Biểu diễn lại tập mẫu và vector trọng số như sau:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \cdots & & & & \\ \cdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix}, w = \begin{pmatrix} w_0 \\ w_1 \\ \cdots \\ w_d \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

- Cần tìm vector tham số w để tối thiểu hàm lỗi:

$$J_n = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^t w)^2 = \frac{1}{n} ||y - Xw||^2.$$

Giải bằng giả nghịch đảo

- Cần tìm vector tham số w để tối thiểu hàm lỗi:

$$J_n = \sum_{i=1}^{n}(y_i - x_i^t w)^2 = ||y - Xw||^2 = (y - Xw)^t(y - Xw)$$

$$\begin{aligned}
&= (y^t - (Xw)^t)(y - Xw) \\
&= (y^t - w^t X^t)(y - Xw) \\
&= y^t y - y^t Xw - w^t X^t y + w^t X^t Xw \\
&= y^t y - y^t Xw - (Xw)^t y + w^t X^t Xw \\
&= y^t y - y^t Xw - y^t (Xw) + w^t X^t Xw \\
&= y^t y - 2y^t Xw + w^t X^t Xw
\end{aligned}$$

- Lấy đạo hàm theo w: $\nabla_w J_n = 2X^tXw - 2X^ty$
- Cho $\nabla_w J_n = 0$ ta có nghiệm: $w^* = (X^tX)^{-1}X^ty$.

Giải bằng giả nghịch đảo

- Nghiệm: $w^* = (X^tX)^{-1}X^ty$, trong đó X là một ma trận $n \times (d+1)$ với các dòng tương ứng với các mẫu và các cột tương ứng với số chiều.
- Ví dụ cho $D = \{(1,2), (2,1),(2,3),(3,2),(4,3),(5,2)\}$.

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 2 \\ 3 \\ 2 \end{bmatrix}$$

- Nghiệm: $w^* = (X^tX)^{-1}X^ty = [1.8615 \quad 0.1077]^t$

Phương pháp gradient descent

- **Mục đích**: tìm vector trọng số w để tối ưu mô hình hồi quy tuyến tính:

$$J_n(w) = Error(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i, w))^2$$

- **Ý tưởng:** Điều chỉnh vector tham số theo hướng giảm của lỗi *Error(w):*

$$w \leftarrow w - \eta \nabla_w Error_i(w)$$

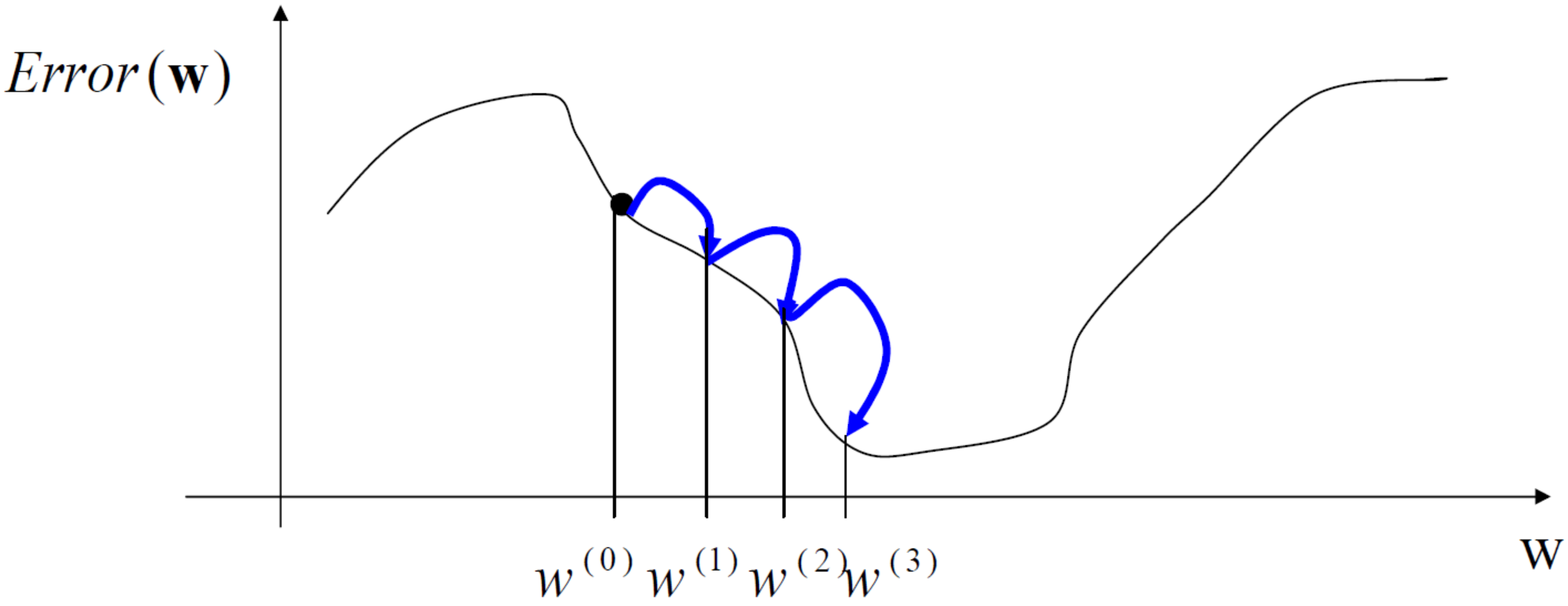
trong đó η là tỷ lệ học (learning rate).

Phương pháp gradient descent

- Điều chỉnh vector tham số theo hướng giảm của lỗi *Error(w):*

$$w \leftarrow w - \eta \nabla_w Error_i(w)$$

- Lặp để tính các giá trị w.



Phương pháp gradient descent

- Mô hình tuyến tính: $f(x) = w^t x$, trong đó $w = (w_0, w_1, \ldots, w_d)^t$ và $x = (1, x_1, \ldots, x_d)^t$.
- Định nghĩa làm lỗi:

$$J(w) = Error_i(w) = \frac{1}{2}(y_i - f(x_i, w))^2 = \frac{1}{2}(y_i - w^t x_i)^2$$

- Tại mỗi bước lặp k:

$$w_j = w_j - \eta(k)\frac{\partial J(w)}{\partial w_j}, \forall j = 1, 2, \cdots, d.$$

tức là:

$$w_j = w_j + \eta(k)(y_i - w^t x_i)x_i, \forall i = 1, 2, \cdots, n, j = 1, 2, \cdots, d.$$

trong đó tốc độ học (learning rate) $\eta(k) = C$ (hằng số) hoặc $\eta(k) = 1/k$.

Phương pháp online gradient

- Mô hình tuyến tính: $f(x) = w^t x$, trong đó $w = (w_0, w_1, \ldots, w_d)^t$ và $x = (1, x_1, \ldots, x_d)^t$.
- Tại mỗi bước lặp:

$$w = w + \eta(k)(y_i - w^t x_i)x_i, \forall i = 1, 2, \cdots, n.$$

- Thuật toán LMS (Least Mean Square):

Algorithm (LMS)

begin initialize $\mathbf{w}$, threshold θ, $\eta(.)$, $k \leftarrow 0$
    do $k \leftarrow (k+1)$
        for $i = 1$ **to** n **do**
            $\mathbf{w} \leftarrow \mathbf{w} + \eta(k)(y_i - \mathbf{w}^t\mathbf{x}_i)\mathbf{x}_i$
    until $|\eta(k)(y_i - \mathbf{w}^t\mathbf{x}_i)\mathbf{x}_i| < \theta$
  return $\mathbf{w}$
end

Phương pháp gradient descent

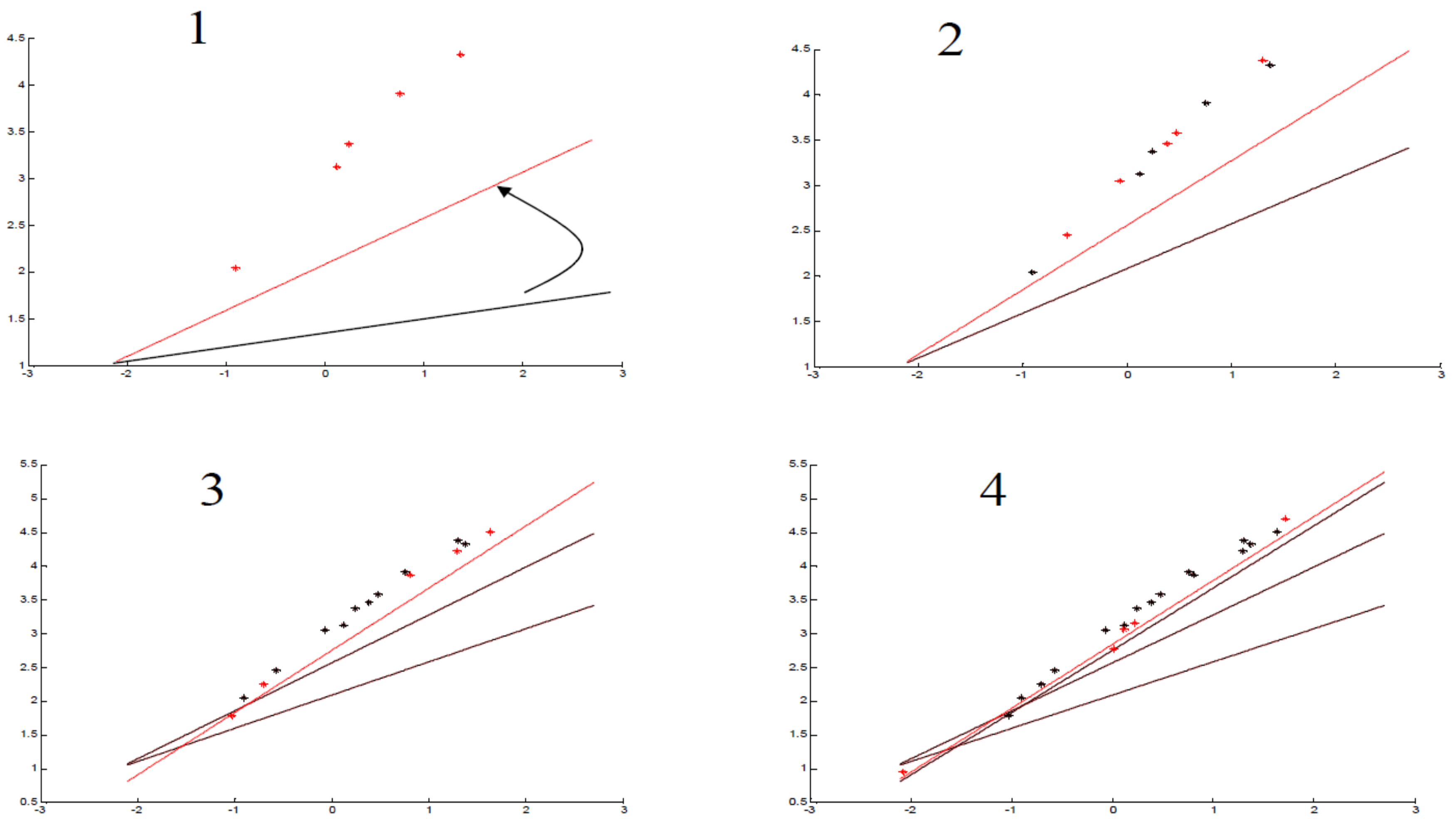
- Ví dụ thuật toán LMS cho dữ liệu: x = [-1, 1, 3, 4, 5] và y = [-5, -1, 3, 5, 7], n = 5 mẫu
- Chuẩn hóa mẫu dữ liệu: $x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 3 & 4 & 5 \end{bmatrix}$, $(d = 2)$
- Khởi tạo: $w = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, θ = 0.1, η = 1, k = 0
- k = 1:
  - $i = 1: w = w + (y_1 - w^t x_1)x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \left(-5 - [0{,}0]\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -5 \\ 5 \end{bmatrix}$
  - $i = 2: w = w + (y_2 - w^t x_2)x_2 = \begin{bmatrix} -5 \\ 5 \end{bmatrix} + \left(-1 - [-5{,}5]\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -6 \\ 4 \end{bmatrix}$
  - $i = 3, w == \begin{bmatrix} -9 \\ -5 \end{bmatrix}; i = 4, w = \begin{bmatrix} 25 \\ 131 \end{bmatrix}, i = 5, w = \begin{bmatrix} -611 \\ -3181 \end{bmatrix}$
  - Khi kết thúc vòng lặp k = 1, lỗi là bao nhiêu?
- k = 2:

Algorithm (LMS)

begin initialize w, threshold θ, $\eta(.)$, $k \leftarrow 0$
    do $k \leftarrow (k+1)$
        for $i = 1$ **to** n **do**
            $\mathbf{w} \leftarrow \mathbf{w} + \eta(k)(y_i - \mathbf{w}^t\mathbf{x}_i)\mathbf{x}_i$
    until $|\eta(k)(y_i - \mathbf{w}^t\mathbf{x}_i)\mathbf{x}_i| < \theta$
  return w
end

Phương pháp online gradient

- Ví dụ



Phương pháp online gradient

- Lập trình bằng Python:
  - Cài đặt Python, ví dụ: C:\Python38
  - Cài đặt thư viện
    - Chuyển thư mục hiện thời vào thư mục ...\Python38\scripts

```
1. pip install numpy
2. pip install matplotlib
3. pip install keras
4. pip install tensorflow-cpu
   pip install tensorflow <hỗ trợ NVIDA>
5. pip install intel-tensorflow
6. pip install pillow
7. pip install opencv-python
```

  - Cài đặt PyCharm

Mở rộng mô hình tuyến tính

- Hàm để học:

$$f(x) = w_0 + \sum_{k=1}^{m} w_k \phi_k(x)$$

- Hàm lỗi:

$$J_n(w) = \frac{1}{n} \sum_{i=1}^{n} (y - f(x_i))^2$$

- Giả sử: $$\varphi(x) = (1, \phi_1(x), \phi_2(x), \cdots, \phi_m(x))^t$$

$$\nabla_w J_n(w) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - f(x_i))\varphi(x_i) = \overline{0}$$

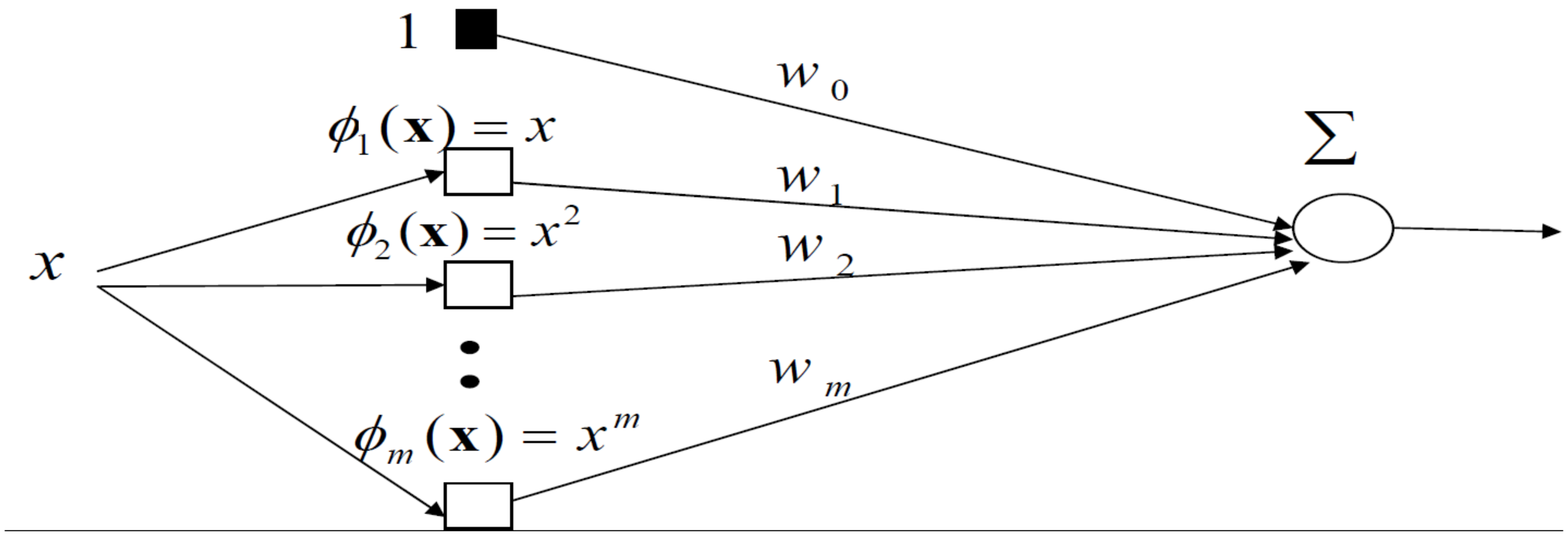
- Dẫn tới một hệ m phương trình tuyến tính:

$$w_0 \sum_{i=1}^{n} 1\phi_j(\mathbf{x}_i) + \ldots + w_j \sum_{i=1}^{n} \phi_j(\mathbf{x}_i)\phi_j(\mathbf{x}_i) + \ldots + w_m \sum_{i=1}^{n} \phi_m(\mathbf{x}_i)\phi_j(\mathbf{x}_i) = \sum_{i=1}^{n} y_i \phi_j(\mathbf{x}_i)$$

Hồi quy với đa thức

- Tập mẫu gồm các bộ (x, y).
- Các hàm cơ sở: $\phi_i(x) = x^i, i = 1, 2, \cdots, m.$
- Hàm để học:

$$f(x, w) = w_0 + \sum_{i=1}^{m} w_i \phi_i(x) = w_0 + \sum_{i=1}^{m} w_i x^i$$



Học với các hàm cơ sở

- Hàm để học:

$$f(x, w) = w_0 + \sum_{i=1}^{m} w_i \phi_i(x)$$

- Cập nhật trực tiếp (online gradient) cho *(x, y):*

$$w_0 = w_0 + \alpha(y - f(\mathbf{x}, \mathbf{w}))$$

$$\vdots$$

$$w_j = w_j + \alpha(y - f(\mathbf{x}, \mathbf{w}))\phi_j(\mathbf{x})$$

Ví dụ hồi quy với đa thức

- Dữ liệu: tập mẫu gồm các bộ (x, y).
- Các hàm cơ sở: $\phi_i(x) = x^i$, $i = 1, 2, \ldots, m$.
- Hàm để học:

$$f(x, w) = w_0 + \sum_{i=1}^{m} w_i \phi_i(x) = w_0 + \sum_{i=1}^{m} w_i x^i = w^t \varphi(x)$$

với $w = (w_0, w_1, w_2, \ldots, w_m)^t$ và $\varphi(x) = (1, x, x^2, \ldots, x^m)^t$.

- Cập nhật trực tiếp (online gradienr) cho bộ (x,y):

$$w_0 = w_0 + \eta(y_i - f(x_i, w)) = w_0 + \eta(y_i - w^t \varphi(x))$$

$$\ldots$$

$$w_j = w_j + \eta(y_i - f(x_i, w))x_i^j = w_j + \eta(y_i - w^t \varphi(x))x_i^j$$

Bài tập

- Cho tập mẫu D = {(-2,4),(-1,1),(0,0),(1,1),(2,4)}. Tìm đa thức bậc 2 để dự đoán cho bộ dữ liệu trên.
- Cho tập mẫu D = {(0,0), (1,4), (2,5), (3,3), (4,2), (5,2), (6,5)}. Tìm đa thức bậc 6 để tối thiểu lỗi trung bình bình phương.