B4. Calculus

Bổ sung cho bài giảng

2019

Nội dung bổ sung

1. **Đạo hàm và tích phân**
2. **Gradient Descent**

1. Đạo hàm và tích phân

❑ **Dãy số** (*sequence*)

$$f : N \to R$$
$$x_n = f(n)$$
$$\{x_n\}_n \equiv \{x_n\} \equiv x_1, x_2, x_3, \cdots, x_n, \cdots$$

❑ **Giới hạn của dãy số (hội tụ)**

$$\lim_{n\to\infty} x_n = a \qquad x_n \to a, n \to \infty$$
$$\forall(\varepsilon > 0), \exists n_0 : \forall n \geq n_0 : |x_n - a| < \varepsilon$$

1. Đạo hàm và tích phân (tt.)

❑ **Một số tính chất cơ bản của giới hạn**

- Mọi dãy hội tụ đều có giới hạn duy nhất
- Mọi dãy hội tụ đều bị chặn
- Giả sử: $\lim_{n\to\infty} x_n = a, \lim_{n\to\infty} y_n = b$. Ta có:

$$(i) \lim_{n\to\infty}(x_n + y_n) = a + b$$
$$(ii) \lim_{n\to\infty}(c + x_n) = c + a$$
$$(iii) \lim_{n\to\infty}(c.x_n) = c.a$$
$$(iv) \lim_{n\to\infty}(x_n.y_n) = a.b$$
$$(v) \lim_{n\to\infty}\left(\frac{x_n}{y_n}\right) = \frac{a}{b}, y_n \neq 0, b \neq 0$$

1. Đạo hàm và tích phân (tt.)

Giới hạn của hàm số

- A là *lân cận* của $x_0 \in \mathbb{R}$: $\exists(\delta > 0): (x_0 - \delta, x_0 + \delta) \subset A$
- Hàm số y = f(x) xác định trên lân cận A của x_0

$$\lim_{x \to x_0} f(x) = L \qquad f(x) \to L, x \to x_0$$

$$\forall(\varepsilon > 0), \exists(\delta > 0):$$
$$\forall |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

Giả sử f(x) xác định trên A. Hàm f(x) liên tục tại $x_0 \in$ A:

$$\lim_{x \to x_0} f(x) = f(x_0)$$

$$\forall(\varepsilon > 0), \exists(\delta > 0):$$
$$\forall |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$$

- Hàm f(x) liên tục trên A nếu f(x) liên tục tại mọi $x \in$ A

1. Đạo hàm và tích phân (tt.)

Một số tính chất cơ bản của đạo hàm

- Nếu f có đạo hàm tại x_0 thì f liên tục tại x_0
- Giả sử $f(x), g(x)$ có đạo hàm tại x. Ta có:

$$(i) \quad (a.f + b.g)'(x) = a.f'(x) + b.g'(x)$$

$$(ii) \quad (f.g)'(x) = f'(x).g(x) + g'(x).f(x)$$

$$(iii) \quad \left(\frac{f}{g}\right)'(x) = \frac{f'(x).g(x) - g'(x).f(x)}{g^2(x)}$$

$$(iv) \quad (f \circ g)'(x) = f'(g(x)).g'(x)$$

1. Đạo hàm và tích phân (tt.)

❑ Đạo hàm của một số hàm sơ cấp

1) $f(x) = a \Rightarrow f'(x) = 0$

2) $f(x) = x \Rightarrow f'(x) = 1$

3) $f(x) = x^{\alpha}, \alpha \in R \setminus \{-1\} \Rightarrow f'(x) = \alpha.x^{(\alpha-1)}$

4) $f(x) = \frac{1}{x} \Rightarrow f'(x) = \frac{-1}{x^2}$

5) $f(x) = a^x, a > 0 \Rightarrow f'(x) = \frac{1}{x.\ln(a)}$

6) $f(x) = e^x \Rightarrow f'(x) = e^x$

7) $f(x) = \log_a(x), a > 0 \Rightarrow f'(x) = a^x.\ln(a)$

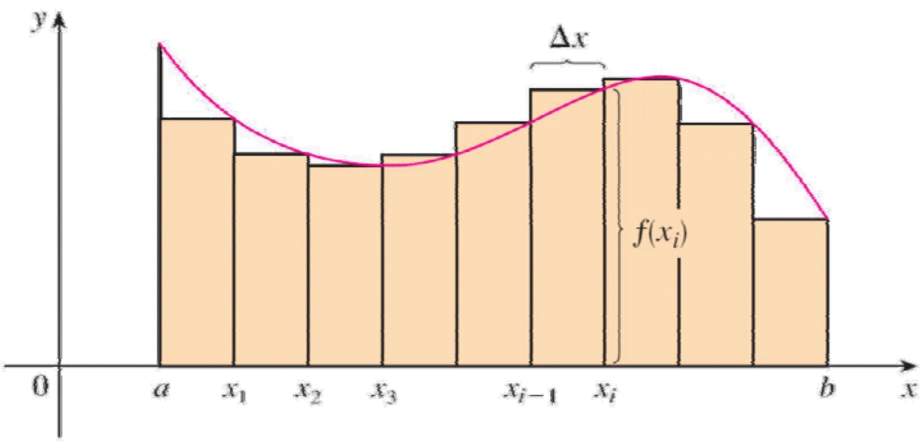
8) $f(x) = \ln(x) \Rightarrow f'(x) = \frac{1}{x}$

9) $f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$

10) $f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$

1. Đạo hàm và tích phân (tt.)

❑ Tích phân xác định

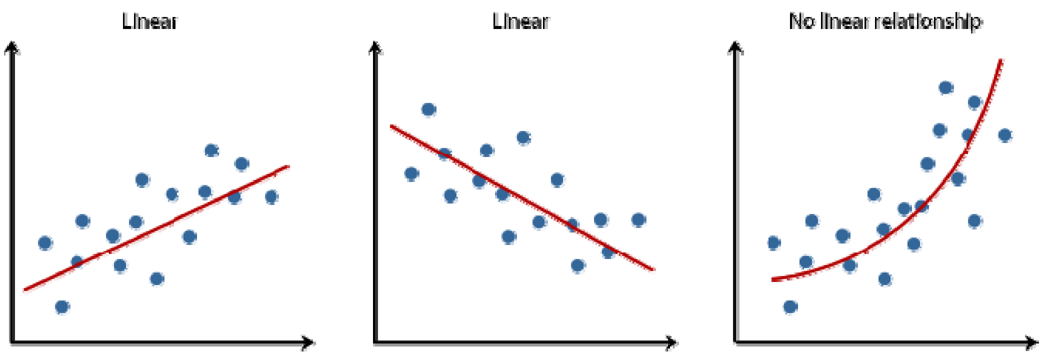


Nội dung bổ sung

1. Đạo hàm và tích phân
2. Gradient Descent

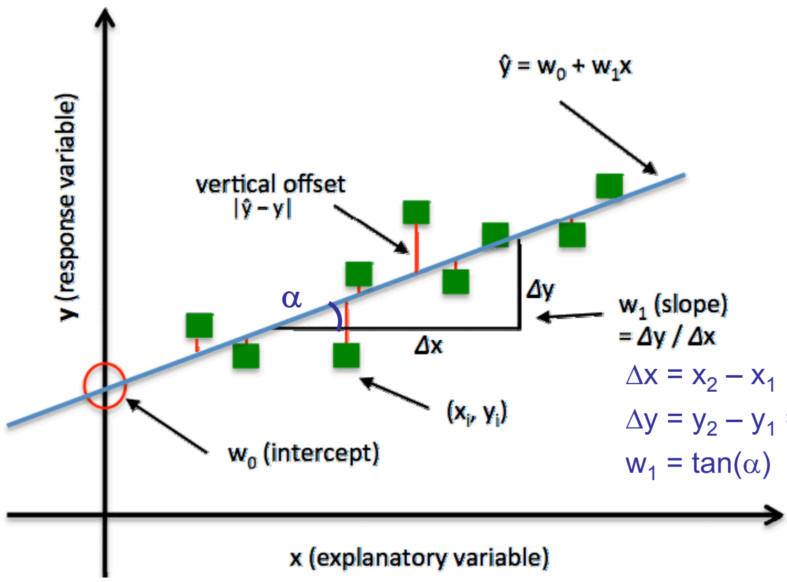
2. Gradient Descent

- **Hồi quy tuyến tính** (*linear regression*)
  - tên khác: *linear fitting*, *linear least square*



2. Gradient Descent (tt.)

❑ **Hồi quy tuyến tính** (*linear regression*)



$\Delta x = x_2 - x_1$

$\Delta y = y_2 - y_1 = f(x_2) - f(x_1)$

$w_1 = \tan(\alpha)$

2. Gradient Descent (tt.)

❑ **Thuật toán Gradient Descent**

- local/global minimum (maximum): $f'(x_0) = 0$
- vòng lặp tìm optimal point x* tiến gần đến x_0 (local minimum)

$f'(x^{(t)}) > 0$: $x^{(t)}$ ở bên PHẢI của $x_0 \Rightarrow$ cần lùi sang TRÁI (A)

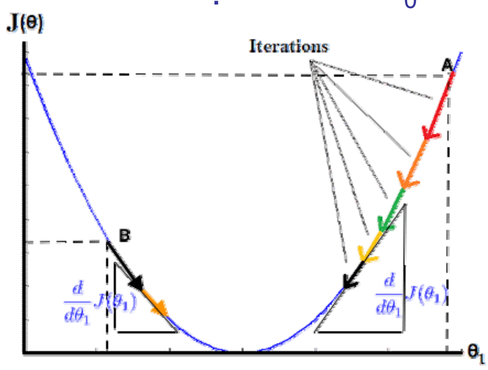
$f'(x^{(t)}) < 0$: $x^{(t)}$ ở bên TRÁI của $x_0 \Rightarrow$ cần tiến sang PHẢI (B)

<u>Tóm lại</u>: $x^{(t)}$ cần di chuyển NGƯỢC DẤU với đạo hàm $\rightarrow x_0$

$$x^{(t+1)} = x^{(t)} - \rho . f'(x^{(t)})$$

$$\theta^{(t+1)} = \theta^{(t)} - \rho . \frac{\partial f(\theta^{(t)})}{\partial \theta^{(t)}}$$

$\rho > 0$: *learning rate* (tốc độ học)



2. Gradient Descent (tt.)

❑ Một số hàm mất mát (loss function)

- *Regression loss*

Mean square error/Quadratic loss/L2 loss: $MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$

Mean absolute error/L1 loss: $MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$

Mean bias error: $MBE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)$

- *Classification loss*

Hinge loss/Multi class SVM loss, Cross entropy loss, …

2. Gradient Descent (tt.)

❑ Sử dụng ma trận giả nghịch đảo

Training set: $T = \{t^{(i)}\}_{i=1}^{m}, t^{(i)} =< x^{(i)}, y^{(i)} >$

input $x^{(i)} \in R^n$ output $y^{(i)} = y_i \in R$

$$X = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \quad Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

bias

Đặt: $x = (x_1, \cdots, x_n) \in R^n, \hat{y} = f(x) = \sum_{j=1}^{n} w_j x_j + w_0$

$$w^T = (w_1, \cdots, w_n, w_0) \in R^{(n+1)}$$

$$\hat{x} = (x_1, \cdots, x_n, \boxed{x_0}) \in R^{(n+1)}, x_0 = 1, \hat{y} = \hat{x}.w$$

$$\hat{x}_i = (x^{(i)}, 1) \in R^{(n+1)}, y_i = y^{(i)}, \forall i = 1, m$$

2. Gradient Descent (tt.)

- **Sử dụng ma trận giả nghịch đảo**

Training set:

$$\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_m \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mn} & 1 \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

Tìm vectơ cột: $w = (w_1, \cdots, w_n, w_0)^T$ sao cho $\hat{y} = \hat{x}.w \approx y$ tốt nhất

Hàm mất mát (*loss function*): $L(w) = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \hat{x}_i w)^2$

Tìm optimal point: $w^* = \arg\min_w L(w)$

2. Gradient Descent (tt.)

- **Sử dụng ma trận giả nghịch đảo**

Hàm mất mát (*loss function*): $L(w) = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \hat{x}_i w)^2$

Tìm optimal point: $w^* = \arg\min_w L(w)$

Xét:

$$\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_m \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mn} & 1 \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

2. Gradient Descent (tt.)

❑ Sử dụng ma trận giả nghịch đảo

$$L(w) = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \hat{x}_i w)^2 \qquad m = |T|$$

- Đạo hàm riêng của L theo w_i

$$C_i$$

$$L(w_i) = (y_i - \hat{x}_i w)^2 = (y_i - \sum_{j=0}^{n}\hat{x}_j w_j)^2 = (y_i - (\hat{x}_i w_i + \sum_{j\neq i}\hat{x}_j w_j))^2 =$$

$$= (y_i - (\hat{x}_i w_i + C_i))^2 = y_i^2 - 2y_i(\hat{x}_i w_i + C_i) + (\hat{x}_i w_i + C_i)^2 =$$

$$= y_i^2 - 2y_i\hat{x}_i w_i - 2y_i C_i + \hat{x}_i^2 w_i^2 + 2\hat{x}_i w_i C_i + C_i^2 =$$

$$= -2y_i\hat{x}_i w_i + \hat{x}_i^2 w_i^2 + 2\hat{x}_i w_i C_i + D_i$$

2. Gradient Descent (tt.)

❑ Sử dụng ma trận giả nghịch đảo

- Đạo hàm riêng của L theo w_i

$$L(w_i) = -2y_i\hat{x}_i w_i + \hat{x}_i^2 w_i^2 + 2\hat{x}_i w_i C_i + D_i$$

$$\frac{\partial L(w_i)}{\partial w_i} = -2y_i\hat{x}_i + 2\hat{x}_i^2 w_i + 2\hat{x}_i C_i = 2\hat{x}_i.(\sum_{j=0}^{n}\hat{x}_j w_j - y_i)$$

$$\frac{\partial L(w)}{\partial w} = \begin{pmatrix} \frac{\partial L(w)}{\partial w_1} \\ \frac{\partial L(w)}{\partial w_2} \\ \vdots \\ \frac{\partial L(w)}{\partial w_n} \end{pmatrix} = \frac{1}{m}\hat{X}^T(\hat{X}.w - Y)$$

2. Gradient Descent (tt.)

- **Sử dụng ma trận giả nghịch đảo**

$$\frac{\partial L(w)}{\partial w} = \frac{1}{m}\hat{X}^T(\hat{X}.w - Y) = 0$$

Giải hệ phương trình, tìm w:

$$\boxed{\underbrace{\hat{X}^T.\hat{X}}_{\mathbf{A}}.w = \underbrace{\hat{X}^T.Y}_{\mathbf{B}}}$$

- Nếu $\hat{X}^T.\hat{X}$ khả nghịch: $w = (\hat{X}^T.\hat{X})^{-1}.\hat{X}^T.Y$
- Nếu $\hat{X}^T.\hat{X}$ KHÔNG khả nghịch: $w = (\hat{X}^T.\hat{X})^{\dagger}.\hat{X}^T.Y$

  với $(\hat{X}^T.\hat{X})^{\dagger}$ là ma trận *giả nghịch đảo* của $\hat{X}^T.\hat{X}$

Tài liệu tham khảo

Vũ Hữu Tiệp, *Machine Learning cơ bản*, 2018