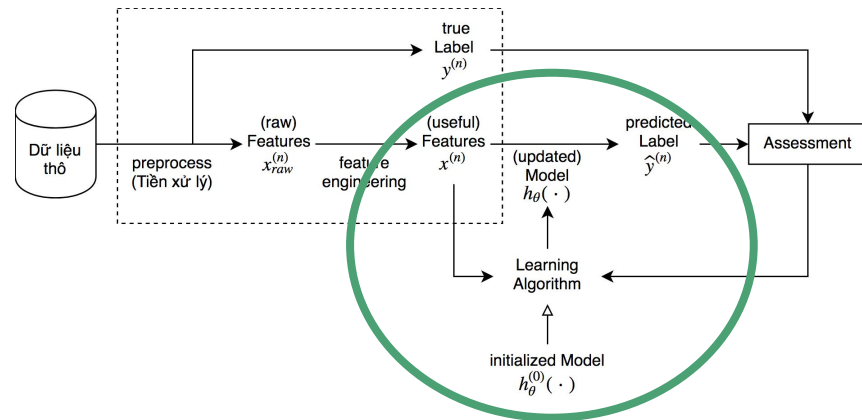


Bài 4: Bài toán Hồi quy & Mô hình Linear Regression

Tuần 2B

Nội dung chính

- Giới thiệu bài toán Hồi quy (Regression)
- Linear regression - một **Mô hình tuyến tính** cho bài toán Hồi quy
 - **Hàm mục tiêu** (objective/cost/loss function)
- Thuật toán học - Learning algorithm
 - Normal equation (closed-form)
 - **Thuật toán Gradient descent** (iterative)



Math notation

- m : số lượng mẫu dữ liệu trong tập huấn luyện $\{(x^{(i)}, y^{(i)}) ; i = 1, \dots, m\}$
- n : số chiều của input vector
- $x^{(i)}$: input vector (**feature vector**) $x^{(i)} \in \mathbb{R}^n$
- X : ma trận với mỗi hàng là input vector
- y : vector chứa các giá trị output tương ứng với các input vectors
- θ : vector trọng số (**model parameter / weight**)
- J : cost function
- ∇J : vector gradient của J
- α : learning rate (tham số phụ trong thuật toán học Gradient Descent)

Giới thiệu bài toán Hồi quy

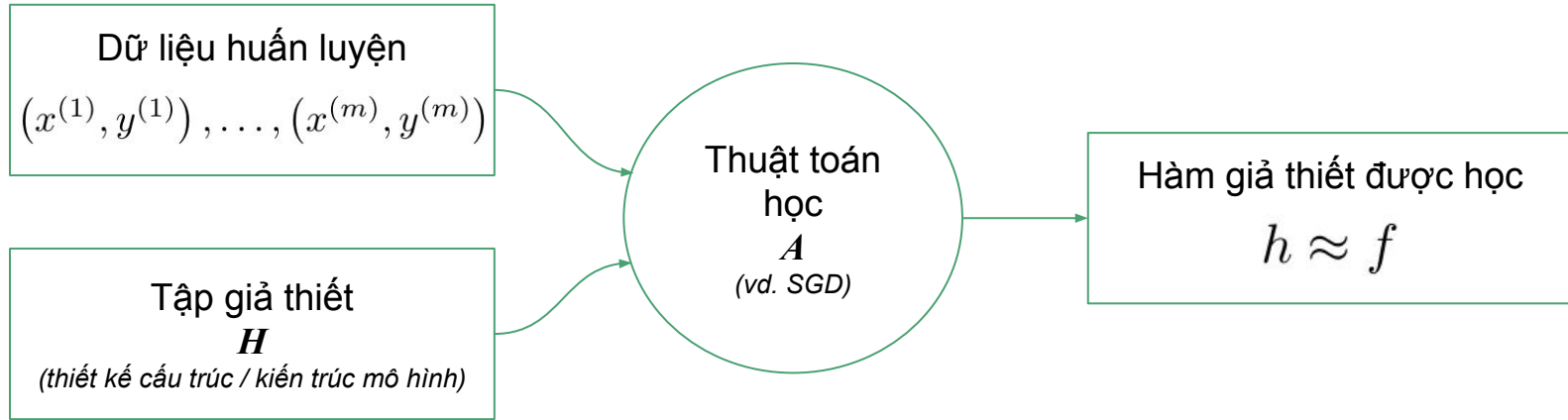
Giới thiệu bài toán Hồi quy

Nhắc lại & Định nghĩa

$$f : X \rightarrow Y$$

(chưa biết, cần tìm)

- Nếu Y là liên tục, bài toán được gọi là **Regression**
- Nếu Y là rời rạc, bài toán được gọi là **Classification**



Giới thiệu bài toán Hồi quy

Ví dụ

Area - x_1	Nr of bedrooms - x_2	Price - y
210	3	400
160	3	330
240	3	369
141	2	232

Given new values $\mathbf{x}_{\text{new}} = [x_1, x_2]$. Estimate y_{new} ?

Our goal is to find \mathbf{h} - unknown hypothesis function. After that: $y_{\text{new}} = h(\mathbf{x}_{\text{new}})$

Giới thiệu bài toán Hồi quy

Ví dụ

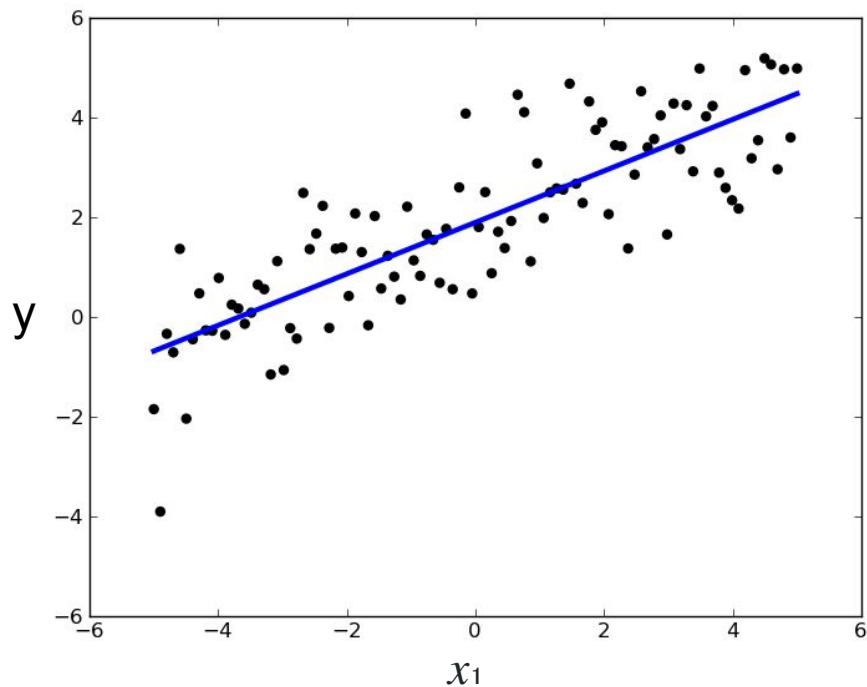
Area - x_1	Nr of bedrooms - x_2	Location - x_3	Price - y
210	3	Quận Tây Hồ	400
160	3	Quận Tây Hồ	330
240	3	Quận Hoàn Kiếm	369
141	2	Quận Cầu Giấy	232

How do you encode x_3 ?

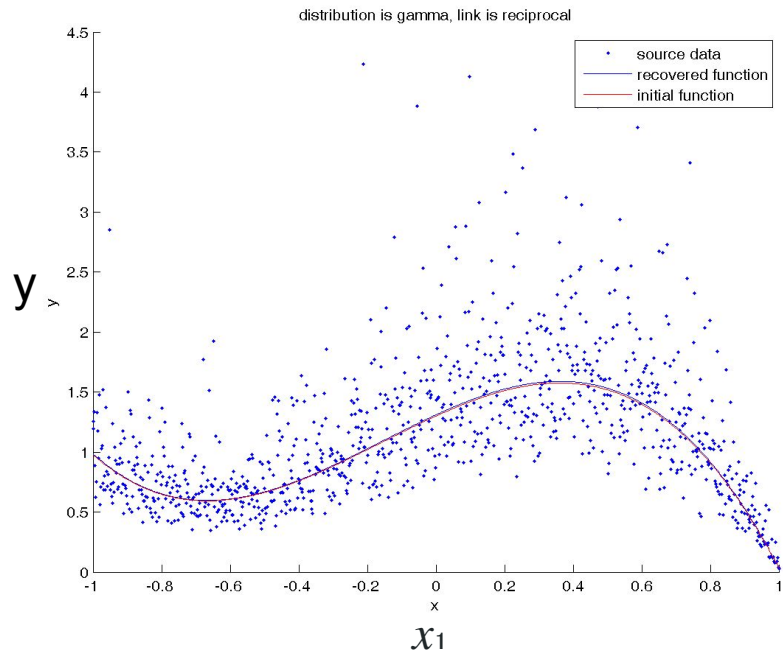
What else need to be done to feature vector \mathbf{x} ?

Mô hình Linear Regression

Thế nào là một Mô hình tuyến tính?



Linear model



Linear model ???

Mô hình Linear Regression

Cấu trúc i.e. dạng của hàm giả thuyết $h(\cdot)$ trong mô hình Linear Regression

Problem

Given a dataset $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$. Assume $x^{(i)} \in \mathbb{R}^1$ for convenience

Build a model that “best fit” i.e. describe your data as close as possible

Model

$$h(\mathbf{x}) = \mathbf{h}_{\theta}(\mathbf{x}) = \theta_0 + \theta_1 x_1 = [\theta_0 \ \theta_1] [1 \ x_1] = \theta^T \mathbf{x}$$

$$h(\mathbf{x}) = \mathbf{h}'_{\theta}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 \sin x_1^2 = [\theta_0 \ \theta_1 \ \theta_2] \begin{bmatrix} 1 & x_1 & \sin x_1^2 \end{bmatrix} = \theta^T \mathbf{x}$$

$\begin{matrix} x_0 & x_1 & x_2 \end{matrix}$

$$h_{\theta}(\mathbf{x}) = \sum_{j=0}^n \theta_j x_j = \boldsymbol{\theta}^T \mathbf{x}$$

- $\mathbf{h}'_{\theta}(\mathbf{x})$ is a **nonlinear** function of data x_1
- $\mathbf{h}'_{\theta}(\mathbf{x})$ is still a **linear** function of unknown parameters $\theta \Rightarrow \mathbf{h}'_{\theta}(\cdot)$ là mô hình tuyến tính (theo θ)

Mô hình Linear Regression

Hàm mục tiêu

$$h_{\theta}(x) = \sum_{j=0}^n \theta_j x_j = \theta^T x$$

Problem

Given a dataset $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$. Assume $x^{(i)} \in \mathbb{R}^1$ for convenience

Build a model that “best fit” i.e. describe your data as close as possible

Learning model

HOW? Find optimal values θ^{best} so that $h_{\theta^{\text{best}}}(x) \approx y$ for as many data points as possible

HOW?? Minimize **objective/cost/loss function**

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2 = \frac{1}{2m} \sum_{i=1}^m \left(\theta^T x^{(i)} - y^{(i)} \right)^2$$

HOW??? (next section)

Thuật toán học cho mô hình Linear Regression

“Thuật toán” Normal Equation (closed-form)

Given a dataset of M samples with N features and M outputs. A datapoint can be represented by a vector of $(N+1)$ dimensions. Then, the dataset can be represented by a matrix of $M * (N+1)$ for inputs and a vector of M dimensions for outputs:

$$x^i = \begin{bmatrix} x_0^i \\ x_1^i \\ \vdots \\ x_n^i \end{bmatrix} \in \mathcal{R}^{N+1} \quad X = \begin{bmatrix} (x^1)^T \\ (x^2)^T \\ \vdots \\ (x^M)^T \end{bmatrix} \in \mathcal{R}^{M*(N+1)} \quad y = \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^M \end{bmatrix} \in \mathcal{R}^M$$

Our cost function is:

$$J(\theta) = \frac{1}{2m} \|X\theta - y\|^2 = \frac{1}{2m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2 = \frac{1}{2m} \sum_{i=1}^m \left(\theta^T x^{(i)} - y^{(i)} \right)^2$$

The gradient of the cost function w.r.t θ is:

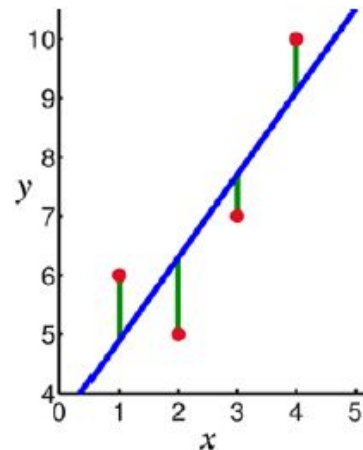
$$\nabla_{\theta} J(\theta) = \frac{1}{m} X^T (X\theta - y)$$

To minimize the cost function, let:

$$\nabla_{\theta} J(\theta) = 0$$

We get the root of the equation

$$\theta = (X^T X)^{-1} X^T y$$



Thuật toán Gradient descent

Giải thuật (iterative)

Khởi tạo ngẫu nhiên tất cả các θ_j

Lặp đến khi hội tụ:

$$\theta_{new} = \theta_{old} - \alpha \nabla J(\theta)$$

(cập nhật tất cả các θ_j đồng thời
trước khi thực hiện iteration tiếp theo)

(α : tham số phụ Learning rate)

Cụ thể

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

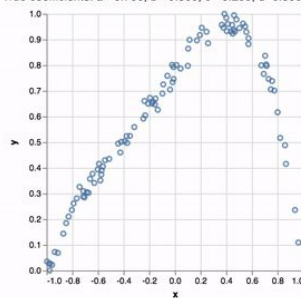
$$\text{vì } \frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^M (h_{\theta}(x^i) - y^i)^2 = \frac{1}{m} \sum_{i=1}^M (h_{\theta}(x^i) - y^i) x_j^i$$

Polynomial Fitting Curve

$$h_{\theta}(x) = ax^4 + bx^3 + cx^2 + dx + e$$

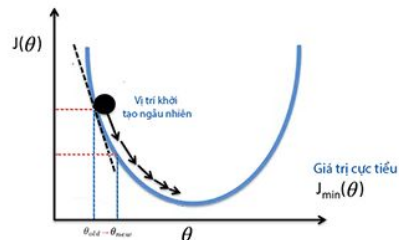
Original Data
True coefficients: a=-0.700, b=-0.800, c=-0.200, d=0.800, e=0.400

Training
Learned coefficients:



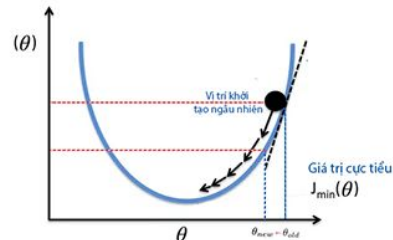
Thuật toán Gradient descent

Tính chất



$$\theta_{new} = \theta_{old} - \alpha \nabla J(\theta)$$

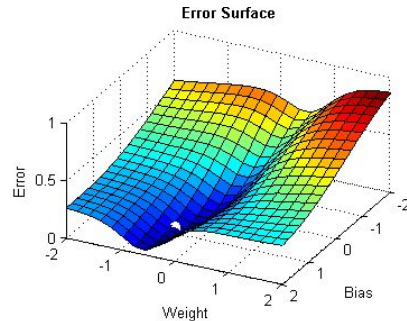
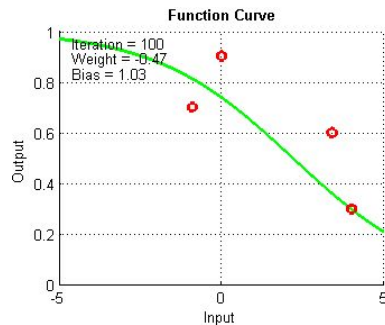
$$\alpha > 0 \\ \nabla J(\theta) < 0 \rightarrow \theta \text{ tăng}$$



$$\theta_{new} = \theta_{old} - \alpha \nabla J(\theta)$$

$$\alpha > 0 \\ \nabla J(\theta) > 0 \rightarrow \theta \text{ giảm}$$

$$h_{\theta}(x) = w \frac{1}{1+e^{-x}} + b$$



Gradient Descent v.s. Normal Equation

Gradient Descent

- Iterative: cần thực hiện nhiều bước lặp để tìm nghiệm tối ưu
- Generic: có thể sử dụng làm thuật toán học **cho nhiều mô hình khác nhau**
- Cần chọn tham số phụ learning rate α , điều kiện hội tụ phù hợp
- Cần lưu ý khi khởi tạo θ_j (vd. trong buổi sau)

Normal Equation

- Closed-form: chỉ cần thực hiện 1 phép tính để tìm nghiệm tối ưu
- Không có tham số phụ
- Khi số đặc trưng (feature) lớn, yêu cầu khả năng tính toán cao do phép nhân $X^T X$ và nghịch đảo có complexity $O(n^2m)$ và $O(n^3)$
- Chỉ sử dụng được làm thuật toán học cho mô hình Linear Regression và số ít các biến thể (vd. Ridge Regression)

Take-home messages

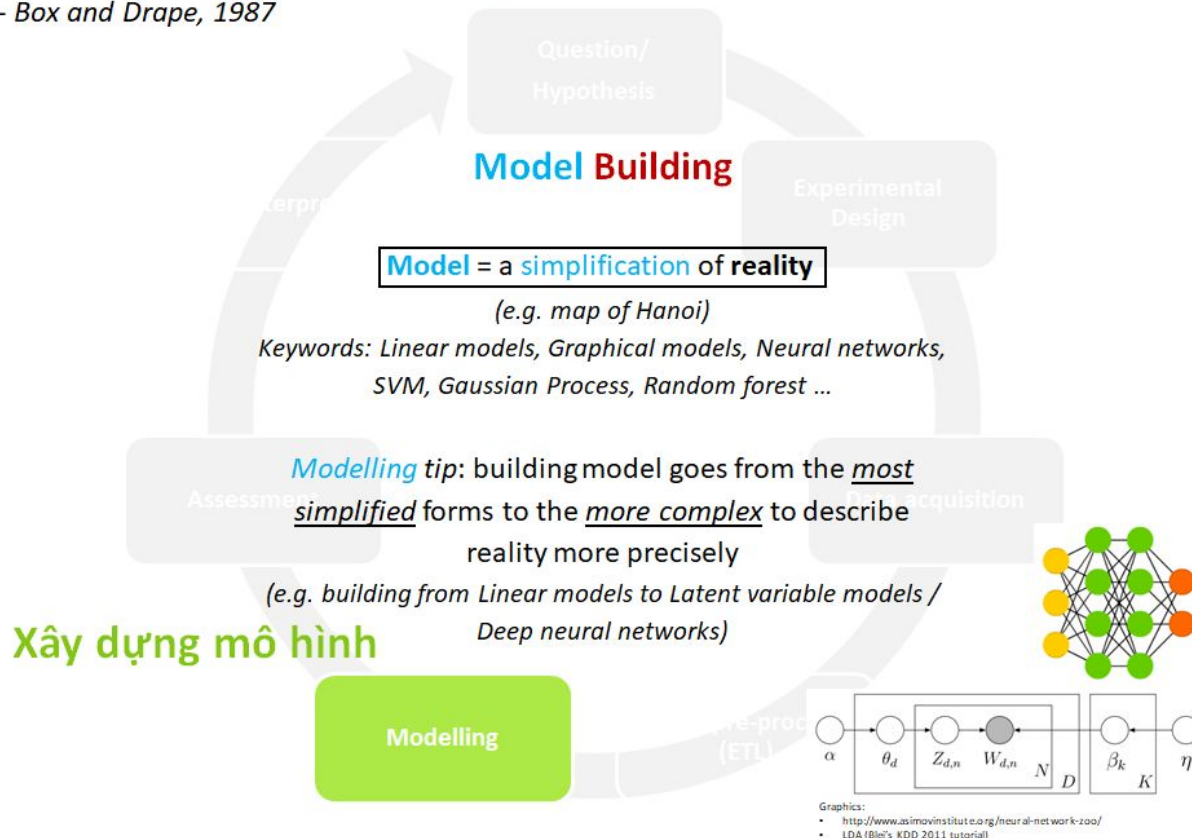
*“The purpose of computation is
insight, not numbers”*

- Richard Hamming

Take-home messages

"All models are wrong, but some are useful."

- Box and Drape, 1987



Slide 7

Area - x_1	Nr of bedrooms - x_2	Location - x_3	Price - y
210	3	Quận Tây Hồ	400
160	3	Quận Tây Hồ	330
240	3	Quận Hoàn Kiếm	369
141	2	Quận Cầu Giấy	232

How do you encode x_3 ?

What else need to be done to feature vector \mathbf{x} ?

Homework

Xây dựng mô hình dự đoán giá nhà ở thành phố Portland, bang Oregon với dữ liệu được cho tại

https://github.com/amaas/stanford_dl_ex/blob/master/ex1/housing.data

Hoặc sử dụng dữ liệu của thành phố Ames, bang Iowa và (tùy chọn) đăng kết quả dự đoán lên Kaggle tại <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview>

Homework

In economics, the Cobb-Douglas functional form of production functions is widely used to represent the relationship of an output to inputs. It was proposed by Knut Wicksell (1851-1926), and tested against statistical evidence by Charles Cobb and Paul Douglas in 1928. In 1928 Charles Cobb and Paul Douglas published a study in which they modeled the growth of the American economy during the period 1899-1922. They considered a simplified view of the economy in which production output is determined by the amount of labor involved and the amount of capital invested.

$$P(L, K) = bL^\alpha K^{1-\alpha}$$

P : total production (the monetary value of all goods produced in a year)

L : labor input (the total number of person-hours worked in a year)

K : capital input (the monetary worth of all machinery, equipment, and buildings)

b : total factor productivity

α : the output elasticities of labor

HW1- Using the data about the American economy in the table to find out b and α

HW2- How to derive the above equation (



award)

Year	P	L	K
1899	100	100	100
1900	101	105	107
1901	112	110	114
1902	122	117	122
1903	124	122	131
1904	122	121	138
1905	143	125	149
1906	152	134	163
1907	151	140	176
1908	126	123	185
1909	155	143	198
1910	159	147	208
1911	153	148	216
1912	177	155	226
1913	184	156	236
1914	169	152	244
1915	189	156	266
1916	225	183	298
1917	227	198	335
1918	223	201	366
1919	218	196	387
1920	231	194	407
1921	179	146	417
1922	240	161	431

Tài liệu tham khảo

1. Linear Regression - Machine Learning Course on Coursera by Andrew Ng.
Alternatively: <http://ufldl.stanford.edu/tutorial/supervised/LinearRegression/>
2. *Chapter 3* - Linear Regression - Pattern Recognition and Machine Learning
C.M.Bishop
3. Machine Learning: Bayesian and Optimization Perspective - Sergios
Theodoridis