Data Analytics

Yong Zheng
Illinois Institute of Technology
Chicago, IL, 60616, USA

School of Applied Technology
ILLINOIS INSTITUTE OF TECHNOLOGY

Schedule

- Coding Practice 2: Classifications
- Exam 2
- Extended Topic: Optimization and Overfitting in Regression Models
- Final Project Notes

Schedule

- Coding Practice 2: Classifications
- Exam 2
- Extended Topic: Optimization and Overfitting in Regression Models
- Final Project Notes

Exam 2

- Time: April 25, 8:35 to 9:50 AM
- Location: SB 111
- Closed books/notes, you can bring a calculator
- For online/remote students, contact IIT online to confirm your exam locations.
- Four questions in total
  - A concept question
  - Manual calculations: KNN and NaiveBayes
  - Read outputs and answer questions: Logistic regression

Schedule

- Coding Practice 2: Classifications
- Exam 2
- Extended Topic: Optimization and Overfitting in Regression Models
- Final Project Notes

Optimizations

Optimizations In Machine Learning

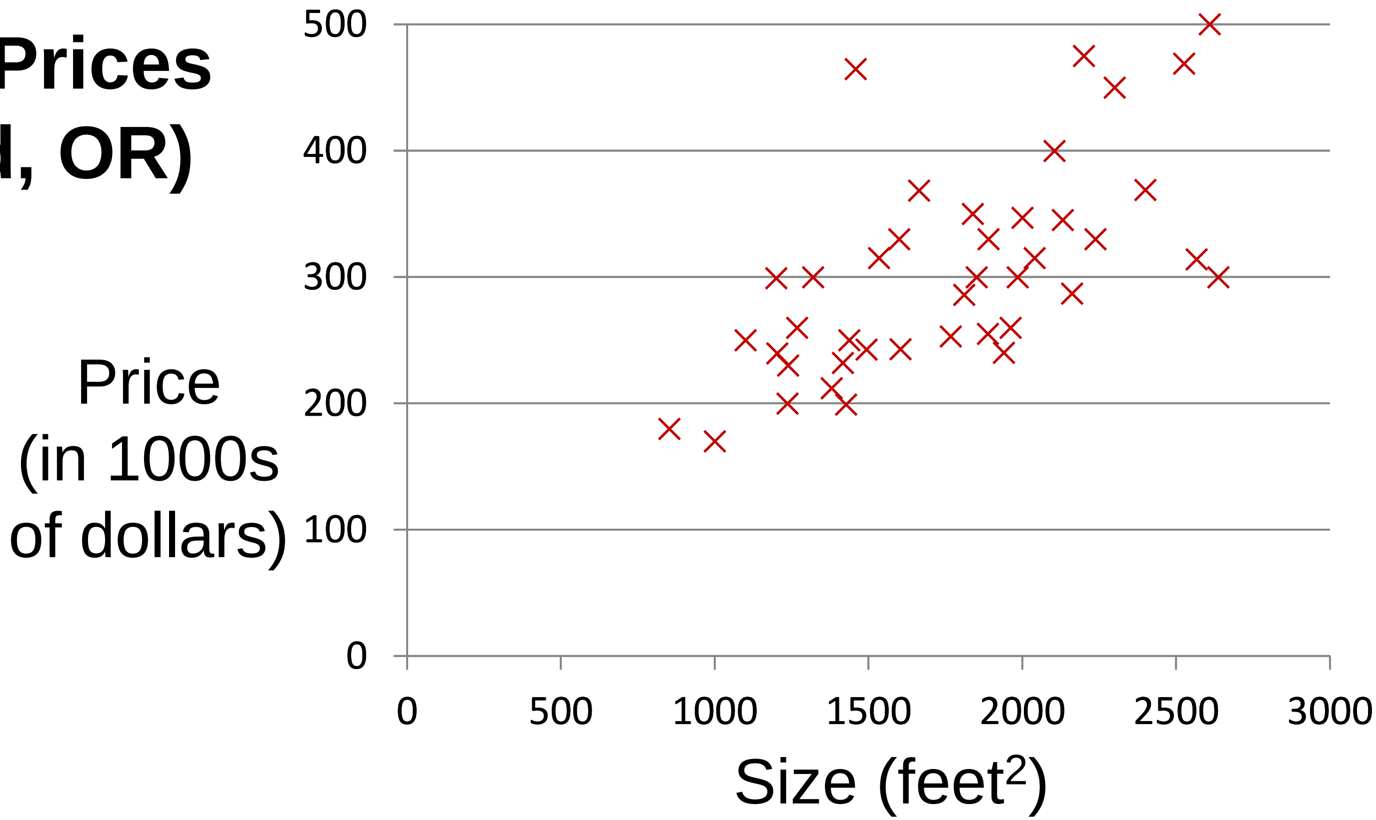
- **Classification or Clustering**
  Divide queries or pages in known groups or groups learned from the data. Examples: adult, news, sports, ...
- **Regression**
  Learn to approximate an existing function. Examples: pulse of a page, stock prices, ...
- **Ranking**
  Not interested in function value but to relative importance of items. Examples: pages or images ranking, ...

Elements In Optimization

- Objective or Cost Function
  - Not only describe your goal. For example, maximize profits or minimize the errors
  - But also formulate and define the objectives formally
- Learning Process or Methods
  - The process could be linear or non-linear
  - It is a convex or concave problem

Example: Linear Regression

Housing Prices (Portland, OR)



Example: Linear Regression

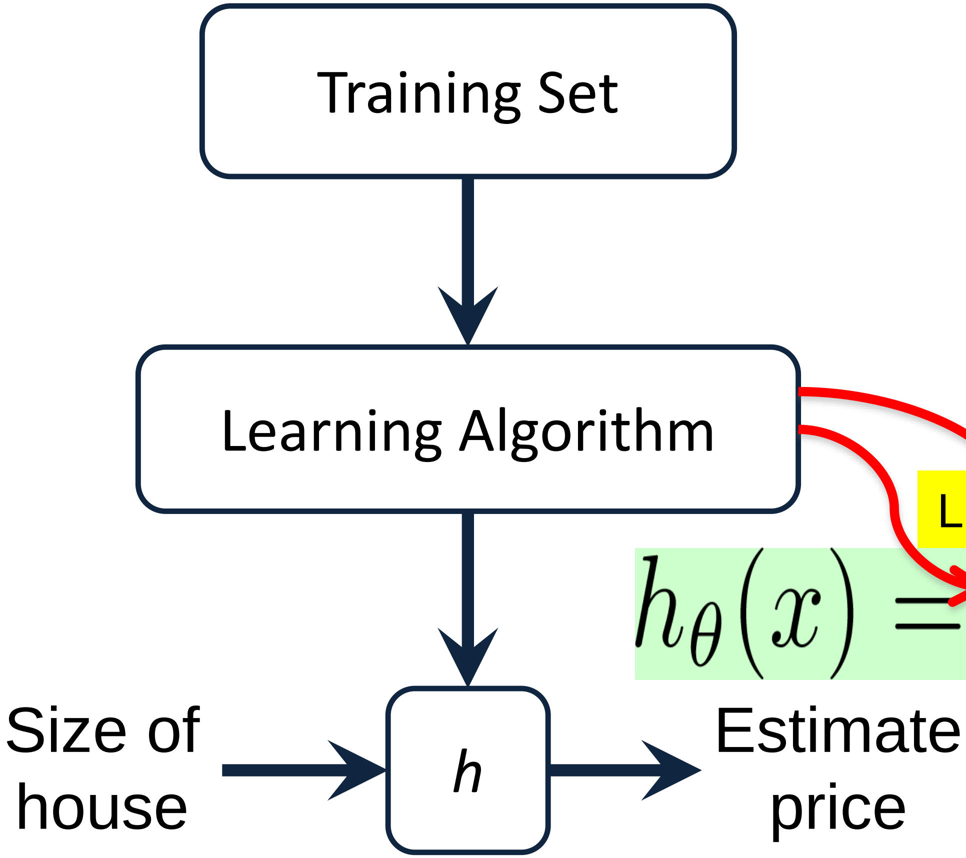
Training set of housing prices (Portland, OR)

Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

Notation:

m = Number of training examples
x' s = "input" variable / features
y' s = "output" variable / "target" variable

Example: Linear Regression



Linear Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Univariate linear regression)

Optimization In Linear Regression

- How to apply gradient descent to minimize the cost function for regression

1. a closer look at the cost function
2. applying gradient descent to find the minimum of the cost function

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function: → Sum of squared errors

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

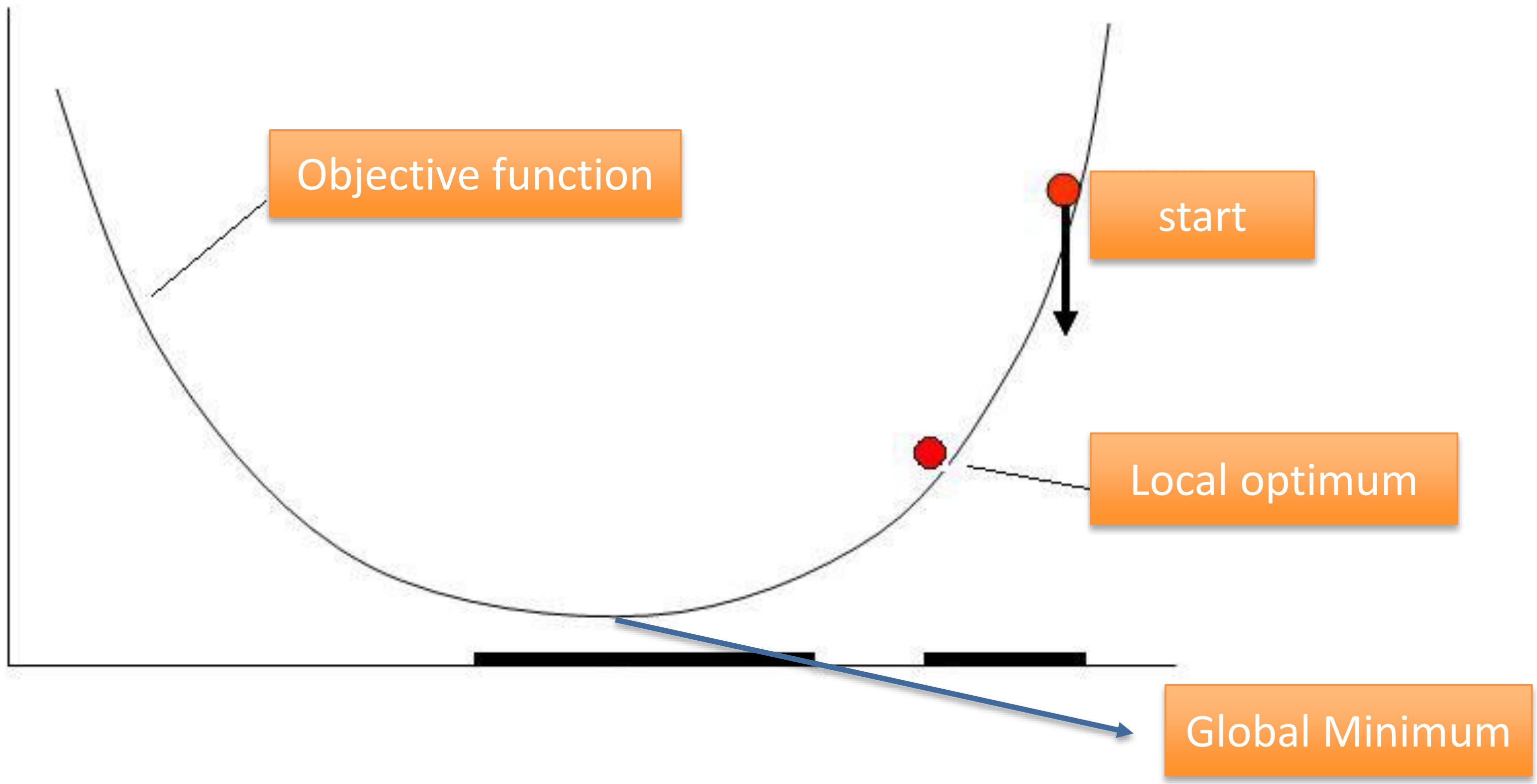
Goal: $\underset{\theta_0, \theta_1}{\text{minimize}} \; J(\theta_0, \theta_1)$

Today

- How to apply gradient descent to minimize the cost function for regression
  1. a closer look at the cost function
  2. applying gradient descent to find the minimum of the cost function

Optimizer: Gradient Descent

- Example of Gradient Descent



Have some function $J(\theta_0, \theta_1)$

Want $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Gradient descent algorithm outline:

- Start with some θ_0, θ_1
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$

  until we hopefully end up at a minimum

Have some function $J(\theta_0, \theta_1)$

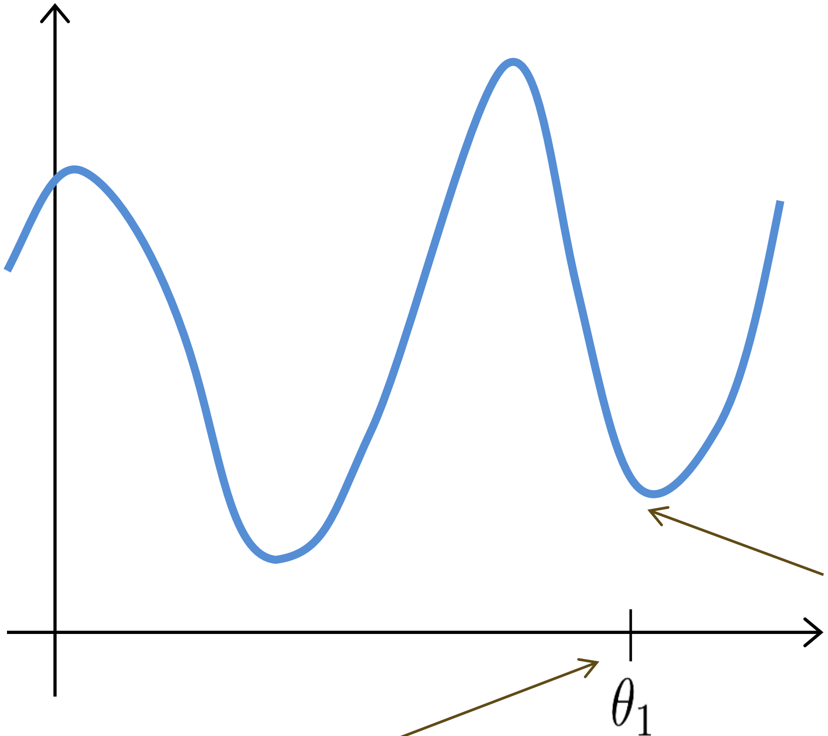
Want $\min\limits_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Gradient descent algorithm

$$
\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{simultaneously update } j = 0 \text{ and } j = 1) \\
&\}
\end{aligned}
$$

Derivative

learning rate



If α is too small, gradient descent can be slow.

If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.

θ_1 at local minimum

Current value of θ_1

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$$(\text{for } j = 1 \text{ and } j = 0)$$

}

Linear Regression Model

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

Gradient descent algorithm

$$
\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \left.\begin{aligned}
\theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \\
\theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x^{(i)}
\end{aligned}\right\} \\
&\}
\end{aligned}
$$

update θ_0 and θ_1 simultaneously

Alleviate Overfittings

Overfitting

- Overfitting is a general issue in all learning process.
- We cannot avoid overfitting, but we can alleviate the issue of overfitting
- How to alleviate overfitting?
  - General Solution: use N-fold cross validation
  - Task-specific Solutions
    - Decision Trees
      - Stop Earlier and Post-prunning
    - Regression Models
      - Ridge and LASSO regression as regularization terms

Overfitting in Regression Models

- We do have the objective function

Cost Function: Sum of squared errors

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

- We want to add *regularization terms* into the cost function
  - Regularization terms is used to apply the penalty in the learning process, in order to alleviate the issue of over-learning!

Overfitting in Regression Models

- How to add the regularization terms?
  - First of all, figure out what are the parameters you are going to learn in the process.
    - In linear regression, we want to learn the intercept and slope. These are the parameters we want to learn
  - Then, decide which regularization term you need
    - L0 term → add a constant value into cost function
    - L1 term → add the absolute value of 1st order terms
    - L2 term → add the 2nd order terms of your parameters

Overfitting in Regression Models

- Example
- Different regularization terms
  - L0 term → add a constant value into cost function
  - L1 term → add the absolute value of 1^{st} order terms
  - L2 term → add the 2^{nd} order terms of your parameters
- Assume our model: $y = \beta_0 + \beta_1 x1 + \beta_2 x2$
- The current cost function = SSE
- New cost function
  - L0 → Cost function = SSE + $\lambda \sum_{n=0}^{N-1} \delta$
  - L1 → Cost function = SSE + $\lambda(|\beta_0| + |\beta_1| + |\beta_1|)$
  - L2 → Cost function = SSE + $\lambda(|\beta_0|^2 + |\beta_1|^2 + |\beta_1|^2)$

L0, L1 and L2

- L0 is specifically designed for sparsity, but the optimization on L0 is a NP-hard problem. We usually use L1 to replace L0
- L1 is LASSO regularization term. It refers to the sum of absolute values of the parameters or a vector. L1 is considered as a feature selection process to make use of the most influential features.
- L2 is Ridge regularization term. It is widely used for alleviating the overfitting problem. It refers to the sqrt of sum of squares of the parameters. L2 is most frequently used one. By default, without special requirements, you should use L2 regularization term

Schedule

- Coding Practice 2: Classifications
- Exam 2
- Extended Topic: Optimization and Overfitting in Regression Models
- Final Project Notes

Final Project

- Each team should have at least one member to present your work.
- Each team has 15 minutes (12 minutes talk + 3 minutes QA)

Final Project

Date	Time	Location	TimeSlot	Group
30-Apr	8:35 - 8:50	SB 212	S1	Explain Exam 2
30-Apr	8:50 - 9:05	SB 212	S2	
30-Apr	9:05 - 9:20	SB 212	S3	245
30-Apr	9:20 - 9:35	SB 212	S4	250
30-Apr	9:35 - 9:50	SB 212	S5	242
2-May	8:35 - 8:50	SB 212	S6	252
2-May	8:50 - 9:05	SB 212	S7	253
2-May	9:05 - 9:20	SB 212	S8	254
2-May	9:20 - 9:35	SB 212	S9	248
2-May	9:35 - 9:50	SB 212	S10	249
6-May	8:00 - 8:15	SB 212	S11	258
6-May	8:15 - 8:30	SB 212	S12	255
6-May	8:30 - 8:45	SB 212	S13	256
6-May	8:45 - 9:00	SB 212	S14	257
6-May	9:00 - 9:15	SB 212	S15	244
6-May	9:15 - 9:30	SB 212	S16	241
6-May	9:30 - 9:45	SB 212	S17	243
6-May	9:45 - 10:00	SB 212	S18	246

Final Projects

- Three steps
  - ~~Project proposal~~
  - Project presentations (10-12 minutes talk + 3-5 minutes QA)
  - Project reports (Due on May 8, 11:59 AM)

Final Projects

- Three steps
  - ~~Project proposal~~
  - Project presentations (10-12 minutes talk + 3-5 minutes QA)
    - The workflow is similar to your final project report, see the template
    - Structure
      - Introduction and motivations
      - Proposed problems
      - Technical solutions, experiments and results
      - Your findings and conclusions
    - Your beginning and ending parts (intro and conclusions) must be easy to be understood by everyone, even if they are non-technical audience
    - At least one member must show up to present your work
  - Project reports (Due on May 8, 11:59 AM)

Final Projects

- Three steps
  - Project proposal
  - Project presentations (10 minutes talk + 5 minutes QA)
  - Project reports (Due on May 8, 11:59 AM)
    - Each team only submits one copy by a single member
    - No extension to the deadline (in noon)
    - What are required to submit
      - Report_Group number.pdf, such as Report_Group 200.pdf
        follow the template to complete the report
      - R Codes_Group number.txt, such as R Codes_Group 200.txt
        provide the R codes only in sequence, provide comments to your codes
      - R Outputs_Group number.pdf, such as R Outputs_Group 200.pdf
        provide the running steps and snapshots for each step, you may paste the codes and also provide the necessary snapshots in this document

Final Projects

- Gradings
  - Overall Quality, 40%
  - Presentation, 15%
  - Codes, 15%
  - Report, 30%
- Feedbacks and comments
  - You will get comments after your presentations
  - You will find your final score for your final project on the blackboard system

Final Projects

- Gradings
  - Overall Quality, 40%
    - Did you meet the basic requirements of the final project
    - How about your experimental design and workflow
    - Did you make any serious mistakes
    - Did you well evaluate and compare your models
    - Did you deliver correct findings, results or conclusions
    - How about your performance in the QA
    - How about your work in comparison with the ones by other teams
    - The degree of ease or difficulty of your projects
  - Presentation, 15%
  - Codes, 15%
  - Report, 30%

Final Projects

- Gradings
  - Overall Quality, 40%
  - Presentation, 15%
    - Did you present well
    - Did you well organize your presentations
    - Did you make sure everyone can understand your intro & conclusions
    - Did you miss any important parts in your presentations
    - How about your performance in the QA
  - Codes, 15%
  - Report, 30%

Final Project: Presentations

- Each team has a total of 15 minutes
  – You should prepare a talk for 10-12 minutes
  – Leave 3-5 minutes for me to give you feedbacks
- You should prepare the laptop with VGA or HDMI port for presentation purpose.
- I may ask you some questions after your presentations
- I will give you feedbacks after each presentation. No more textual comments/feedbacks on the blackboard systems

Final Project: Presentations

- You only have one chance to present
- No 2nd chance for you/your team if you had some problems during the presentation

General Guides To A Good Presentation

- Some other factors you may want to know
  - Visualization? Make sure it is clear.
    Not too small, not too large
  - Less texts, more figures/visualizations
    The audience usually do not want you to read the texts on the slides!
  - Do NOT exceed the time

Final Projects

- Gradings
  - Overall Quality, 40%
  - Presentation, 15%
  - Codes, 15%
    - Did you correctly submit the necessary documents
    - Did you provide clear and neat coding with necessary comments
    - Did you provide clear and correct outputs/snapshots
  - Report, 30%

Final Projects

- Gradings
  - Overall Quality, 40%
  - Presentation, 15%
  - Codes, 15%
  - Report, 30%
    - Is your report clear and correct
    - Especially, can your solutions solve the proposed problems
    - Did you provide right and clear experiments and results
    - Did you deliver right findings and conclusions
    - Did you fix the problems given based on my comments after your presentations