

Introduction

CS 229: Machine Learning

Emily Fox

Stanford University

January 8, 2024

©2024 Emily Fox

1

AI is hot!

CS 229: Machine Learning

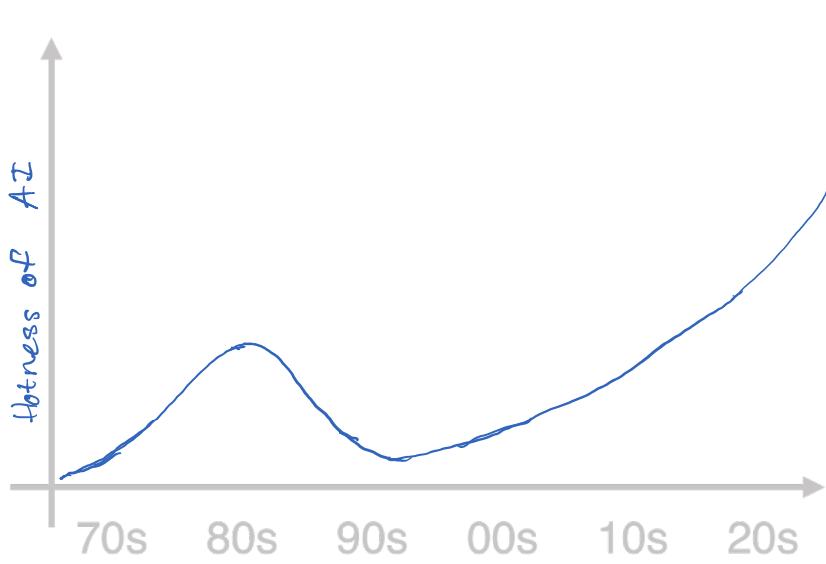
2

1

Resurgence of AI

CS 229: Machine Learning

3



CS 229: Machine Learning

4

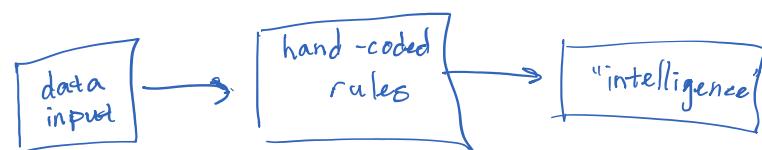
Trend 1: Don't code, learn

©2024 Emily Fox

CS 229: Machine Learning

5

Most AIs in the 80s were hand-coded...

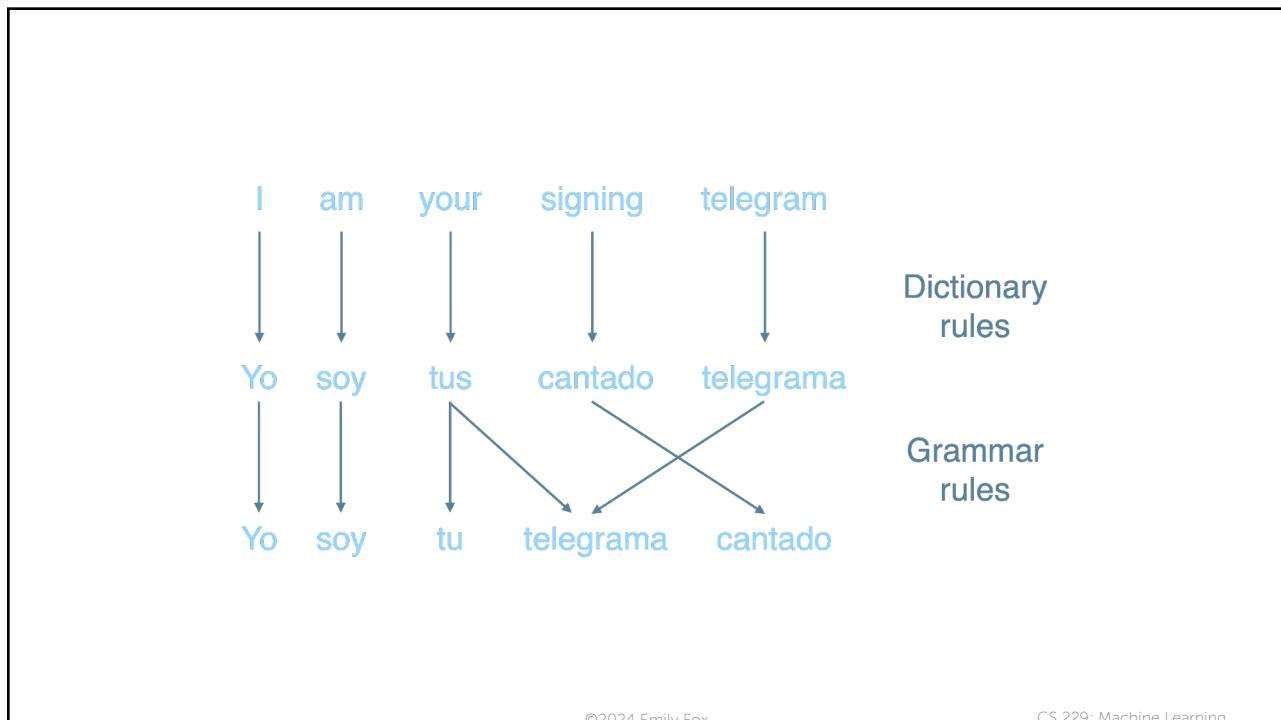


©2024 Emily Fox

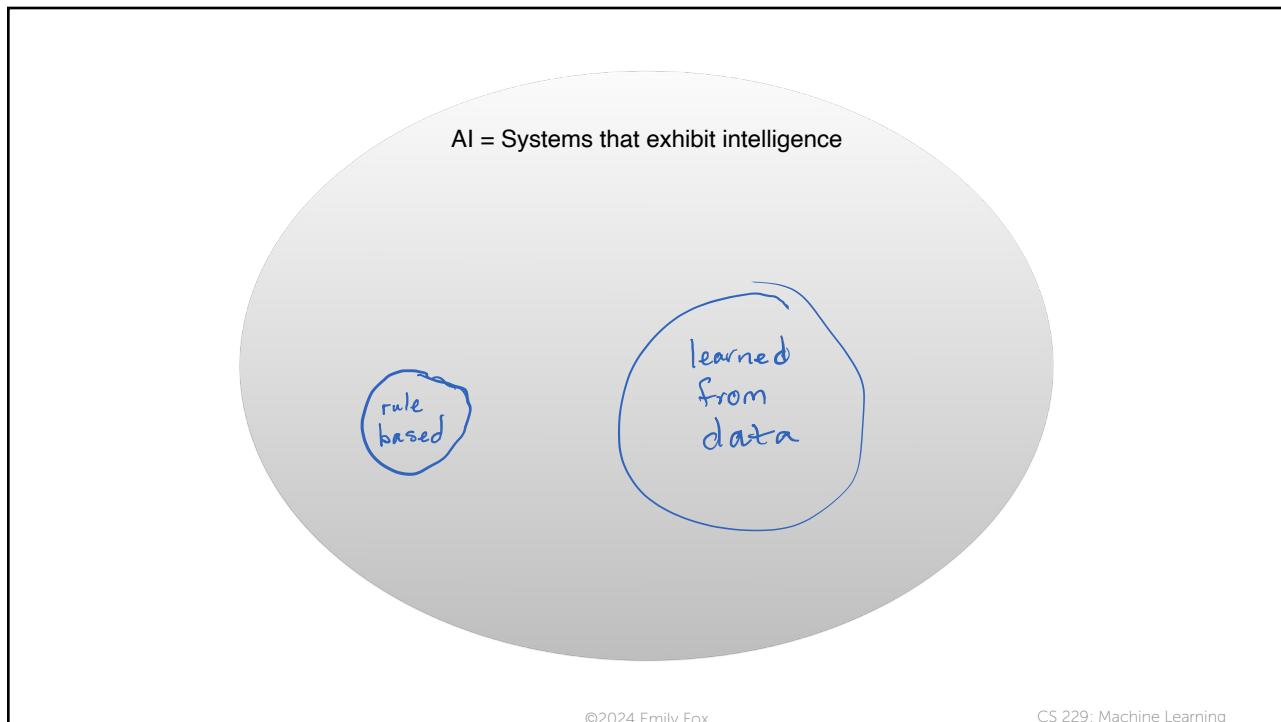
CS 229: Machine Learning

6

3



7



8

Flashback to 1957...

Rosenblatt's Perceptron



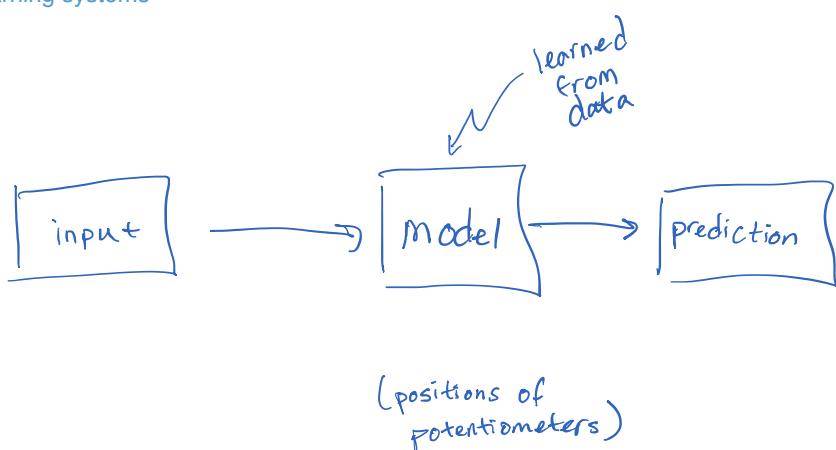
©2024 Emily Fox

CS 229: Machine Learning

9

Today, best AIs are learned

Machine learning systems



©2024 Emily Fox

CS 229: Machine Learning

10

Flashback to 1998...

©2024 Emily Fox

CS 229: Machine Learning

11

Deep Blue vs AlphaGo



Deep Blue: Defeats Kasparov, Chess Grand Master - IBM 1997



AlphaGo: Defeats Sedol, 9 dan Go Player - Google 2016

©2024 Emily Fox

CS 229: Machine Learning

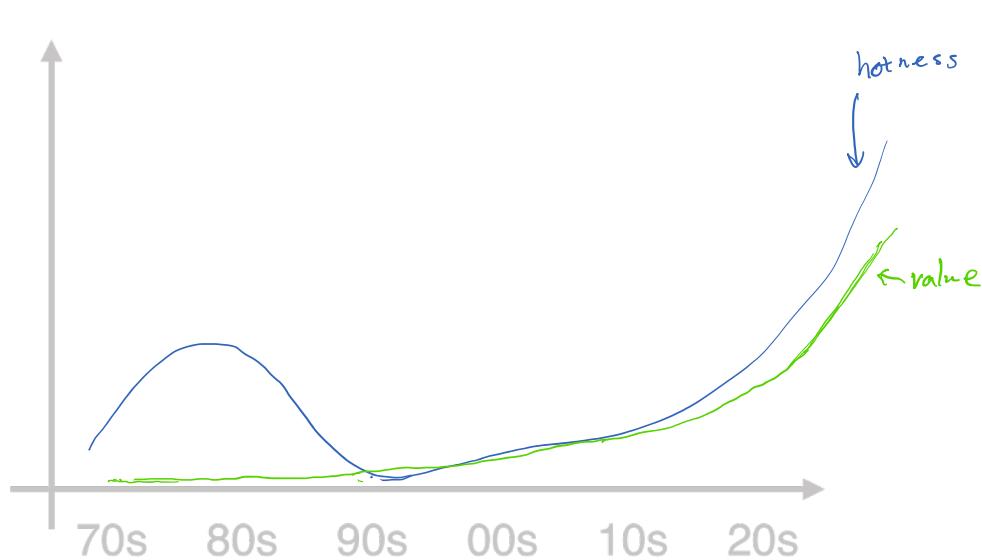
12

What's driving the change?

©2024 Emily Fox

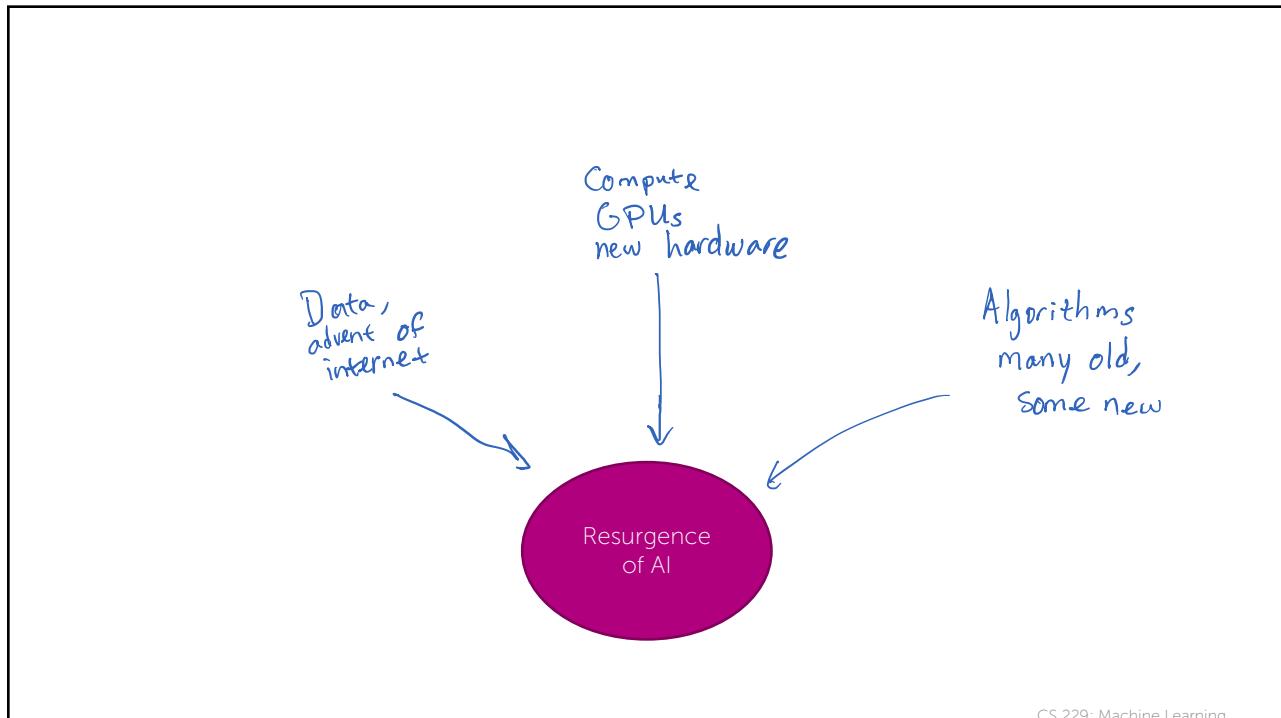
CS 229: Machine Learning

13



CS 229: Machine Learning

14

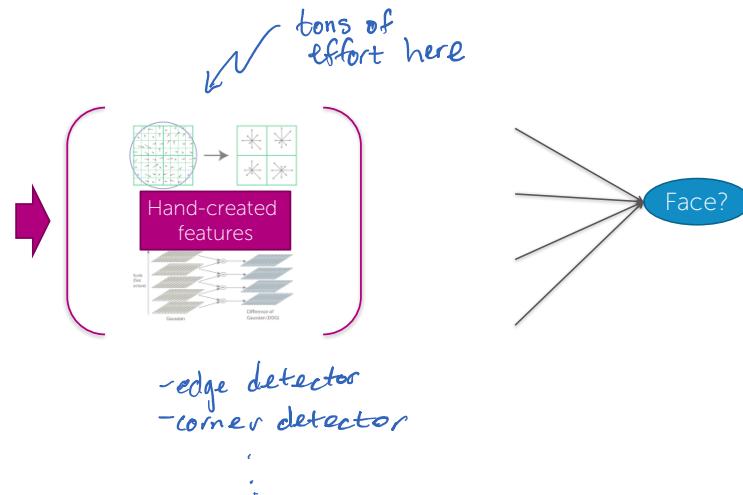


15

Trend 2: Deep learning fever

16

Standard machine learning pipeline

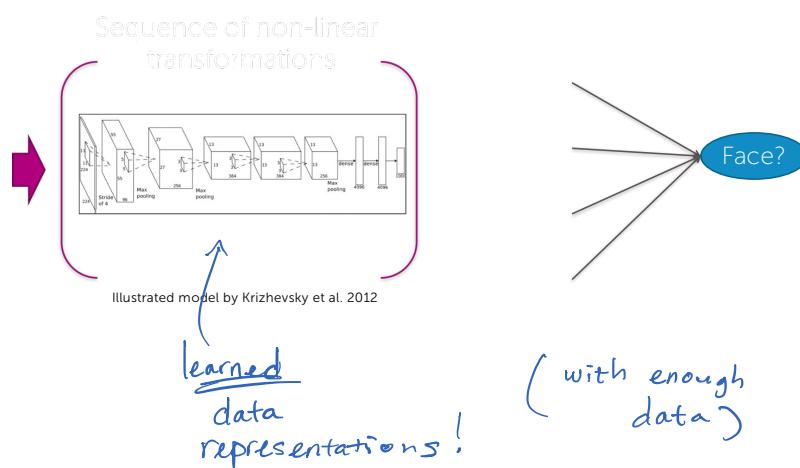


©2024 Emily Fox

CS 229: Machine Learning

17

Deep learning pipeline

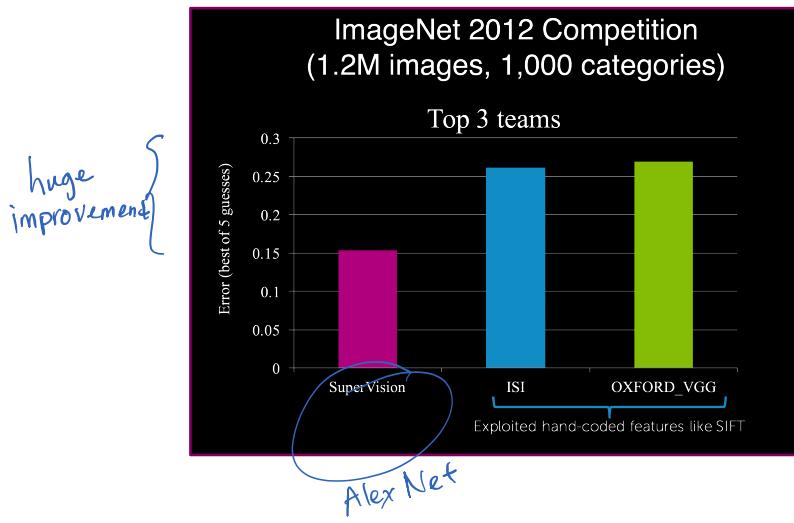


©2024 Emily Fox

CS 229: Machine Learning

18

Why are we feverish about deep learning?

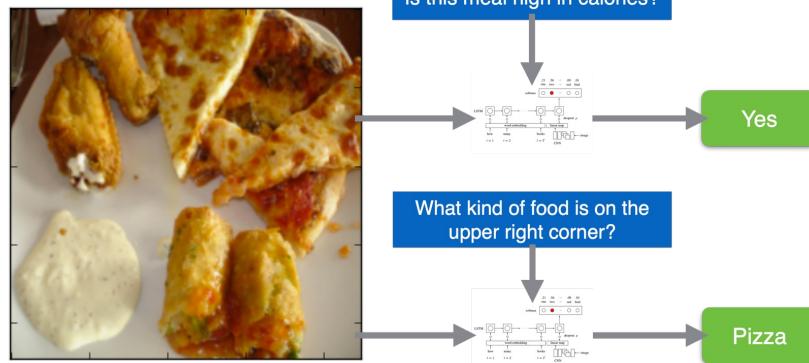


19

Deep learning is very “modular”...
Image deep features + text deep features + deep classifier

20

Visual question answering



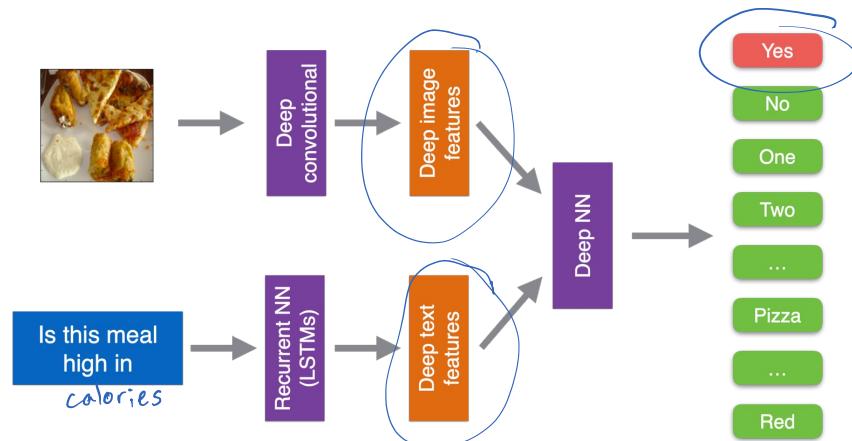
©2024 Emily Fox

CS 229: Machine Learning

21

Deep learning models are very modular...

E.g., Ren et al 2015 for VQA



©2024 Emily Fox

CS 229: Machine Learning

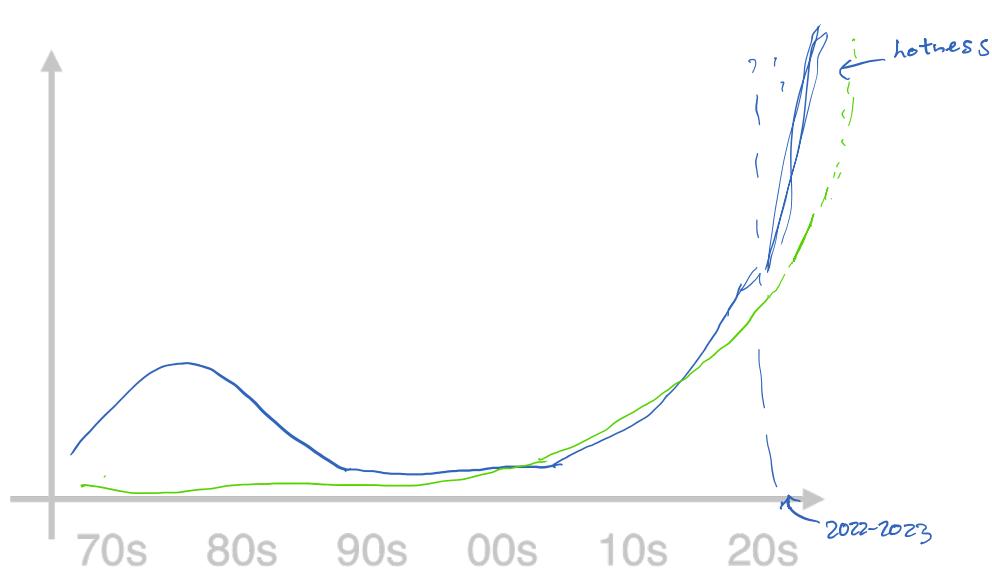
22

Trend 3: Large language models are bringing AI to everyone

©2024 Emily Fox

CS 229: Machine Learning

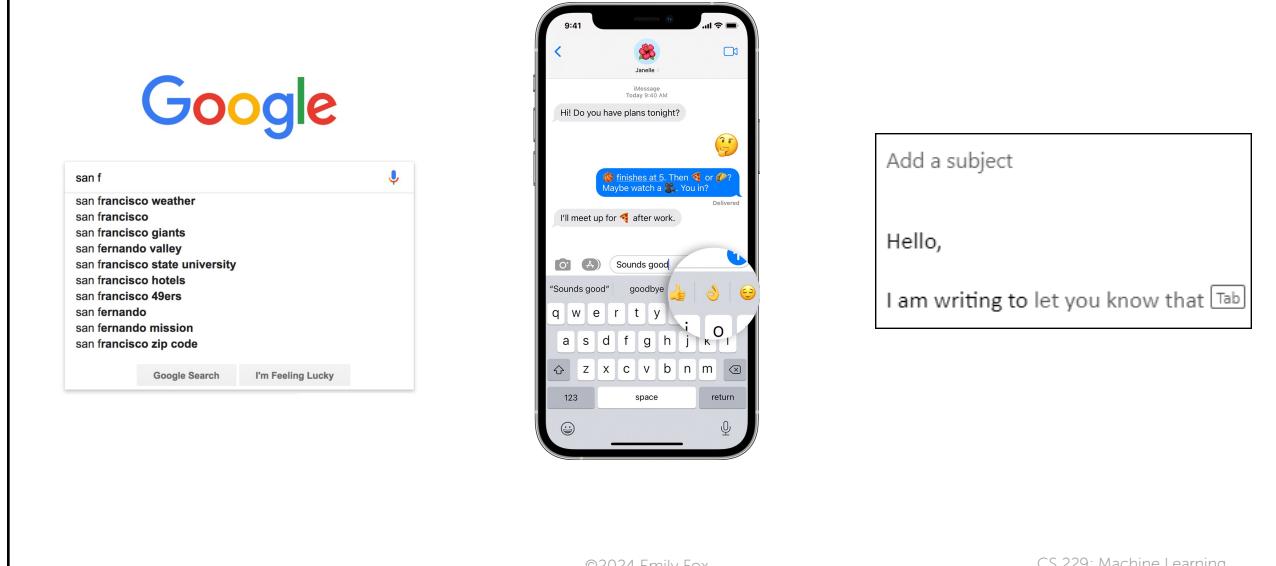
23



CS 229: Machine Learning

24

Language models for autocomplete

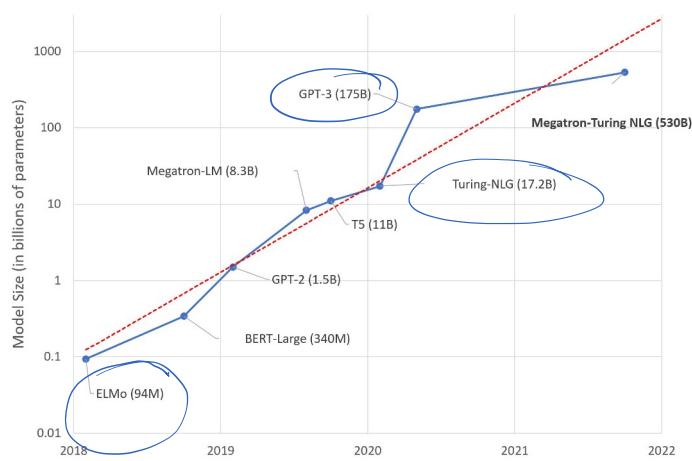


©2024 Emily Fox

CS 229: Machine Learning

25

Language models have been getting bigger...



©2024 Emily Fox

CS 229: Machine Learning

26

When language models get big enough, new capabilities start to emerge...

@2024 Emily Fox

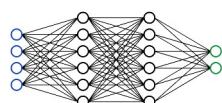
CS 229: Machine Learning

27

Foundation models: emergence

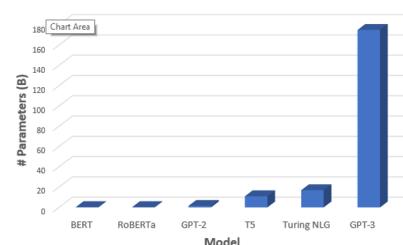
self-supervised learning

In 1885, Stanford _____
In 1885, Stanford University was _____



+

scale



= **emergence**

Find a word that rhymes: duck, luck; lunch, **munch**

CS 229: Machine Learning

28

Prompt

Summarize this for a second-grade student:

Jupiter is the fifth planet from the Sun and the largest in the Solar System. It is a gas giant with a mass one-thousandth that of the Sun, but two-and-a-half times that of all the other planets in the Solar System combined. Jupiter is one of the brightest objects visible to the naked eye in the night sky, and has been known to ancient civilizations since before recorded history. It is named after the Roman god Jupiter. [19] When viewed from Earth, Jupiter can be bright enough for its reflected light to cast visible shadows.[20] and is on average the third-brightest natural object in the night sky after the Moon and Venus.

Sample response

Jupiter is a planet that is bigger than all the other planets in our solar system and is very bright when you see it in the night sky. It is named after the Roman god Jupiter. When viewed from Earth, it is usually one of the three brightest objects in the sky.

Prompt

Create an outline for an essay about Nikola Tesla and his contributions to technology:

1. Introduce Nikola Tesla and his work in the field of electricity and magnetism.
 2. Discuss Tesla's contributions to the development of alternating current (AC) electricity.
 3. Describe Tesla's development of the Tesla coil and its impact on wireless technology.
 4. Discuss Tesla's experiments with X-rays and their implications for the future of medical technology.
 5. Conclude with a discussion of Tesla's legacy and his impact on the development of modern technology.

Prompt

Convert movie titles into emoji.

Back to the Future: 🎉🍿🎥
 Batman: 🦇🦇🦇
 Transformers: 🚗🤖🤖
 Star Wars: 🌐⭐️

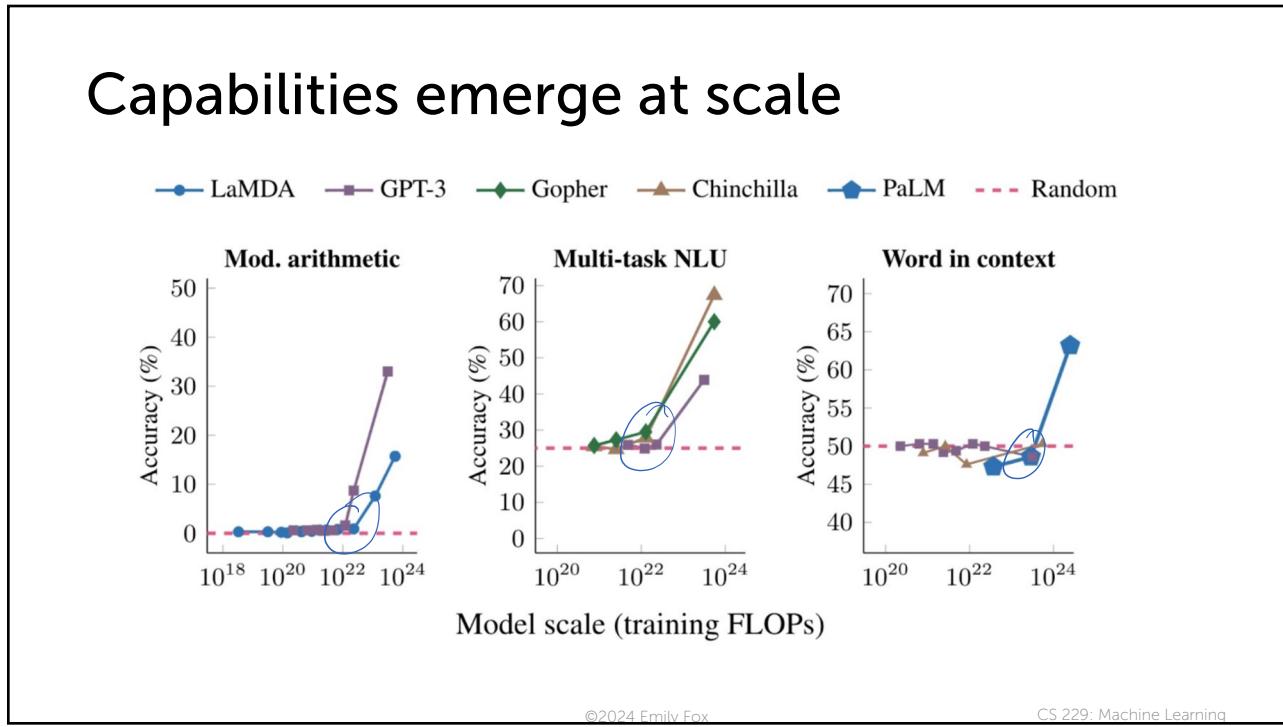
Sample response

🌟

OpenAI's GPT-3

©2024 Emily Fox CS 229: Machine Learning

29



30

AI Translations (AI Output in Blue)

Legal clause: Upon liquidation of the Company, the Series A Shareholders will receive in preference to all other shareholders an amount in respect of each Series A Share equal to one times the Original Issue Price (the "Liquidation Preference"), plus all accrued but unpaid dividends. To the extent that the Company has assets remaining after the distribution of that amount, the Series A Shareholders will participate with the holders of Ordinary Shares pro rata to the number of shares held on an as converted basis.

Plain English: If the startup is wound up, the Series A investors will be paid back at least what they invested and they will also share any leftover assets with ordinary shareholders.

Legal Clause: Sale of all or substantially all of the assets of the Company or a sale of shares involving a change in control (each, a "Corporate Transaction") will be treated in the same way as a liquidation and the proceeds of sale will be distributed as set out in paragraph 3. If the holders of Series A Shares have received any Special Dividend it shall be set off against their Liquidation Preference.

Plain English: If the company is sold, or a new owner takes control, the proceeds of the sale will be distributed as in the liquidation clause above. Any special dividend paid will be treated as an initial payment towards the Series A investors.

©2024 Emily Fox

CS 229: Machine Learning

31

Describe a layout.

Just describe any layout you want, and it'll try to render below!

a button that looks like a watermelon Generate

```
<button style={{backgroundColor: 'pink', border: '2px solid green', borderRadius: '50%', padding: 20, width: 100, height: 100}}>Watermelon</button>
```



©2024 Emily Fox

CS 229: Machine Learning

32

Code from comments

```

1 import datetime
2
3 def parse_expenses(expenses_string):
4     """Parse the list of expenses and return the list of triples (date, value, currency).
5     Ignore lines starting with #.
6     Parse the date using datetime.
7     Example expenses_string:
8         2016-01-02 -34.01 USD
9         2016-01-03 2.59 DKK
10        2016-01-03 -2.72 EUR
11    """
12     expenses = []
13     for line in expenses_string.splitlines():
14         if line.startswith("#"):
15             continue
16         date, value, currency = line.split(" ")
17         expenses.append((datetime.datetime.strptime(date, "%Y-%m-%d"),
18                          float(value),
19                          currency))
20
21     return expenses

```

 Copilot

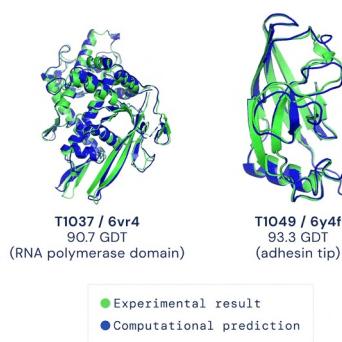
GitHub CoPilot (powered by OpenAI's Codex)

©2024 Emily Fox

CS 229: Machine Learning

33

Protein folding



DeepMind's AlphaFold, UW's RoseTTAFold, Meta's ESMFold

©2024 Emily Fox

CS 229: Machine Learning

34

Image generation

©2024 Emily Fox

CS 229: Machine Learning

35

GANs [Goodfellow et al. 2014]



©2024 Emily Fox

CS 229: Machine Learning

36

Generating images from text

Examples generated with midjourney



pirate ship in the sea with a pirate kid smiling, children's book illustration, modern, naif, colorful, luminous, Lisa Wee by @franpaezgrillo



Lonely tree Forgotten night sky, 4K, high quality by @apslq



a person riding a bicycle fast down a hill, 4k by @guestrin

©2024 Emily Fox

CS 229: Machine Learning

37

Where are we heading?

©2024 Emily Fox

CS 229: Machine Learning

38

19

AI will reshape our social norms
and the very essence of how we
interact with each other

CS 229: Machine Learning

39

Trend 4: A quest for a more ethical AI

©2024 Emily Fox

CS 229: Machine Learning

40

20

Language models for autocomplete

Google search results for "san f":

- san f
- san francisco weather
- san francisco
- san francisco giants
- san fernando valley
- san francisco state university
- san francisco hotels
- san francisco 49ers
- san fernando
- san fernando mission
- san francisco zip code

iPhone messaging screen:

Hi! Do you have plans tonight? 🤔

I'll meet up for 🍷 after work. Sounds good! Goodbye! 😊

Add a subject

Hello,

I am writing to let you know that [Tab]

©2024 Emily Fox

CS 229: Machine Learning

41

Who's voice is being amplified?

Send

To: Basically Everyone

Cc:

Re: Your Email

I am sorry for the delayed response [Tab]

©2024 Emily Fox

CS 229: Machine Learning

42

Are Emily and Greg More Employable than Lakisha and Jamal? [Bertrand & Mullainathan '03]



©2024 Emily Fox

CS 229: Machine Learning

43

ML-based system for recruiting

- Could decrease this bias...
- But, could also amplify biases...

©2024 Emily Fox

CS 229: Machine Learning

44

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Machine Learning

45

Science

Current Issue First release papers Archive About Submit manuscript

RESEARCH ARTICLE

Dissecting racial bias in an algorithm used to manage the health of populations

ZIAD OBERMEYER, BRIAN POWERS, CHRISTINE VOGELI AND SENDHIL MULLAINATHAN Authors Info & Affiliations

SCIENCE • 25 Oct 2019 • Vol. 366, Issue 6464 • pp. 447–453 • DOI:10.1126/science.aay2342

MIT News
ON CAMPUS AND AROUND THE WORLD

SUBSCRIBE

How machine-learning models can amplify inequities in medical diagnosis and treatment

MIT researchers investigate the causes of health care disparities among underrepresented groups.

Steve Nadis | MIT CSAIL
August 17, 2023

ACLU

About Issues Our work News Take action Shop Give

NEWS & COMMENTARY

Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism

Unclear regulation and a lack of transparency increase the risk that AI and algorithmic tools that exacerbate racial biases will be used in medical settings.

©2024 Emily Fox

CS 229: Machine Learning

46

What CS229 is all about

©2024 Emily Fox

CS 229: Machine Learning

47

**This is not a state-of-the-art
deep learning models course**

©2024 Emily Fox

CS 229: Machine Learning

48

24

**In this course, you will learn the
fundamentals of ML,
so you can successfully and correctly apply ML in
practice and be ready to develop new methods**

©2024 Emily Fox

CS 229: Machine Learning

49

We will cover...

- Various learning settings
 - Supervised, unsupervised, self-supervised, reinforcement learning,...
- Key modeling approaches
 - Linear models, neural nets, decision trees, boosting, clustering, transformers,...
- Main tools
 - Optimization, EM, regularization, bias-variance tradeoff,...
- Ethical boundaries
 - Explainability, algorithmic bias, privacy,...
- And more...

50

©2024 Emily Fox

CS 229: Machine Learning

50

25

**It's a fast-paced course...
you'll need to stay on top of the work,
but you will learn to
be fluent in machine learning**

©2024 Emily Fox

CS 229: Machine Learning

51

Course logistics

©2024 Emily Fox

CS 229: Machine Learning

52

Course Staff

INSTRUCTORS



Emily Fox



Sanmi Koyejo

COURSE STAFF



Zhoujie D. (Head TA)



John C. (Course Manager)



Rishi A.



Samir A.



Rishi D.



Kefon D.



Hongjun J.



Jacob F.



Sonia C.



Winnie C.



Alex W.



John S.



Ryan S.



Hermann K.



Mark X.

©2024 Emily Fox

CS 229: Machine Learning

53

53

Supportive Community Values

Please Reach Out if You Need Help

- We encourage all members of CS229 to reach out whenever they need assistance or guidance.
- Your success and well-being are important to us, and we are here to support you.

Community Over Individual

- In CS229, we firmly believe that as a community, we are greater than the sum of our parts.
- Together, we can achieve more and overcome challenges that may seem insurmountable individually.

Embrace Empathy

- Practicing empathy is a fundamental part of our community values.
- Understand the perspectives and challenges of your fellow students and colleagues.
- Be willing to lend a helping hand when someone is in need.

Helping Others Matters

- The act of helping others not only benefits them but strengthens our community.
- Small gestures of assistance can have a significant impact on someone's journey.

54

©2024 Emily Fox

CS 229: Machine Learning

54

Primary Communication Platform: ED

- Why ED?

- **Efficiency:** Utilizing Ed for inquiries enhances the efficiency of our communication channels. It provides a centralized platform to engage in discussions, share information, and seek clarification.
- **Timely Responses:** Ed allows for prompt responses to inquiries, ensuring that students receive the assistance they need precisely when they need it..
- **Course Updates:** Important course updates, announcements, and clarifications will be posted on Ed.

Guidelines for Effective Use:

- **Be Specific:** When posting inquiries, please be as specific as possible to facilitate accurate responses.
- **Respectful Interaction:** Maintain a respectful and constructive tone in all communications to create a positive learning environment.
- **Check Frequently:** Make it a habit to check Ed regularly for updates, announcements, and responses to inquiries.

Encouraging Peer Engagement:

- **Collaborative Learning:** We highly encourage students to answer each other's questions on Ed. Collaborative learning enriches the educational experience for everyone involved.
- **Incentive:** To incentivize this collaboration, we will be giving **2% extra credit** of the total course grade to the ten students with the most instructor-endorsed responses on Ed by the end of the quarter.

55

@2024 Emily Fox

CS 229: Machine Learning

55

Course Assessment Overview

Four Assignments (40%)

Midterm Examination (20%)

Final Project & Poster Session (40%)

- Open-ended
- Teams of 1-3 (4 only in special cases)

Weighted Assignment Grading

- The assignments are weighted based on their respective point values.
- To calculate the weight of each assignment, we use the following formula:
 - If $\{p1, p2, p3, p4\}$ denotes the point values of each assignment, then the weight of HW1 is calculated as:
 - $HW1\ Weight = (p1 / (p1 + p2 + p3 + p4)) * 40\%$

Grading Basis

- This quarter's grading basis offers two options: letter grade or Credit/No Credit (CR/NC).
- It is your responsibility to ensure that your desired grading basis is correctly enrolled on Axess.

56

@2024 Emily Fox

CS 229: Machine Learning

56

Course Assessment Overview (continued)

- PS1 due: January 24th
- PS2 due: February 7th
- Midterm: February 16th
- PS3 due: February 21st
- PS4 due: March 6th
- Final Project Reports due: March 15th
- Poster Session: March 20th (3:30-6:30pm)

57

@2024 Emily Fox

CS 229: Machine Learning

57

Collaboration Policy

Study Groups and Accountability

- Collaboration within study groups is encouraged for problem-solving.
- Students must independently write and understand solutions without relying on joint notes.

Honor Code Awareness & Guidelines

- Familiarize yourself with the Stanford Honor Code and its relevance to CS courses and refer to our Course Logistics for CS229.
- If uncertain about violations, create an Ed post or contact the Honor Code Lead.
- Copying, referring to, or comparing solutions from previous years or online sources is strictly prohibited.
- If seeking help or discussing problems with others, proper citation in your submission is mandatory.

Online Posting and Plagiarism Detection

- Do not post assignment solutions online or share them on public repositories.
- Plagiarism-detection software compares your work against past submissions and solutions.

AI Tools Policy

- You may use generative AI tools such as Co-Pilot and ChatGPT as you would use a human collaborator.
- You may not directly ask generative AI tools for answers or copy solutions.
- The use of generative AI tools to substantially complete an assignment or exam (e.g., by directly copying) is prohibited and will result in honor code violations. We will be checking students' homework to enforce this policy.
 - For more details: Generative AI reference [here](#).

58

@2024 Emily Fox

CS 229: Machine Learning

58

Prerequisites:

- Knowledge of basic computer science principles and skills, at a level sufficient to write a reasonably non-trivial computer program in Python/numpy. (CS106A or CS106B, CS106X.)
- Familiarity with probability theory. (CS 109, MATH151, or STATS 116)
- Familiarity with multivariable calculus and linear algebra (relevant classes include, but not limited to MATH 51, MATH 104, MATH 113, CME 205, CME 100.)

Note: These prerequisite skills are essential for success in this course and will be assumed throughout the curriculum. Please ensure you have the necessary background knowledge before proceeding.

59

@2024 Emily Fox

CS 229: Machine Learning

59

Friday TA Lectures

- These supplementary lectures, led by Teaching Assistants (TAs), serve as a valuable resource for students to reinforce their understanding of prerequisite concepts and enhance their grasp of the lecture material.
- The sessions will be conducted on a weekly basis, specifically every Friday.
- While attendance is not mandatory, we strongly recommend students to participate, as it can significantly contribute to their overall comprehension and success in the course.
- The first of these supplementary sessions is scheduled for this week on January 12th. We look forward to seeing you there!

60

@2024 Emily Fox

CS 229: Machine Learning

60

AI is changing the world.

**You are now taking a big step in
being part of this transformation**

©2024 Emily Fox

CS 229: Machine Learning

61

Regression

CS 229: Machine Learning
Emily Fox
Stanford University
January 8, 2024

©2024 Emily Fox

62

Regression

©2024 Emily Fox

CS 229: Machine Learning

63

How much is my house worth?



64

©2024 Emily Fox

CS 229: Machine Learning

64

How much is my house worth?



65

©2024 Emily Fox

CS 229: Machine Learning

65

Data



input *output*
 $(x_1 = \text{sq.ft.}, y_1 = \$)$



$(x_2 = \text{sq.ft.}, y_2 = \$)$



$(x_3 = \text{sq.ft.}, y_3 = \$)$



$(x_4 = \text{sq.ft.}, y_4 = \$)$



$(x_5 = \text{sq.ft.}, y_5 = \$)$

⋮

Input vs. Output:

- y is the quantity of interest
- assume y can be predicted from x

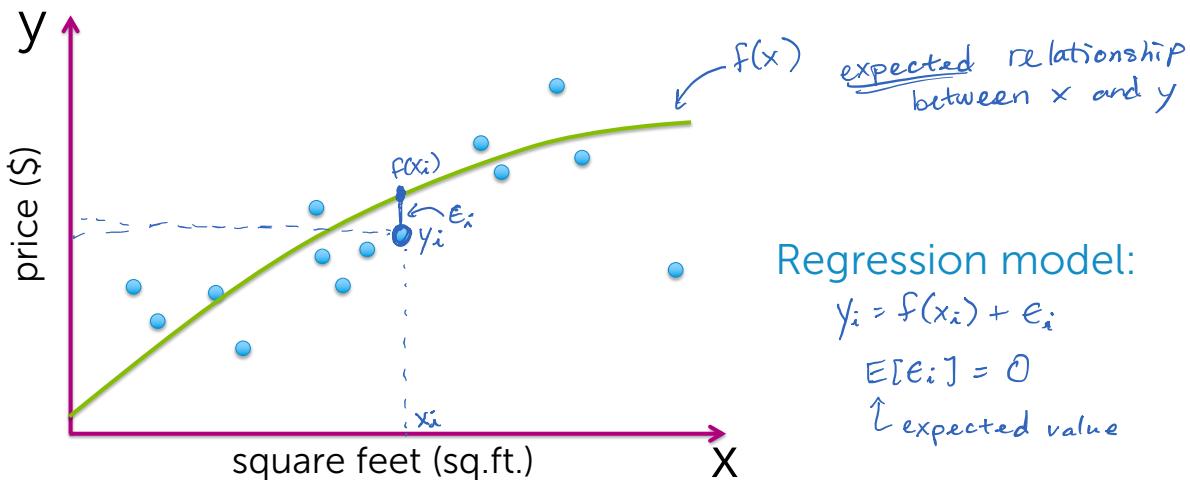
Example of supervised learning

©2024 Emily Fox

CS 229: Machine Learning

66

Model – How we assume the world works



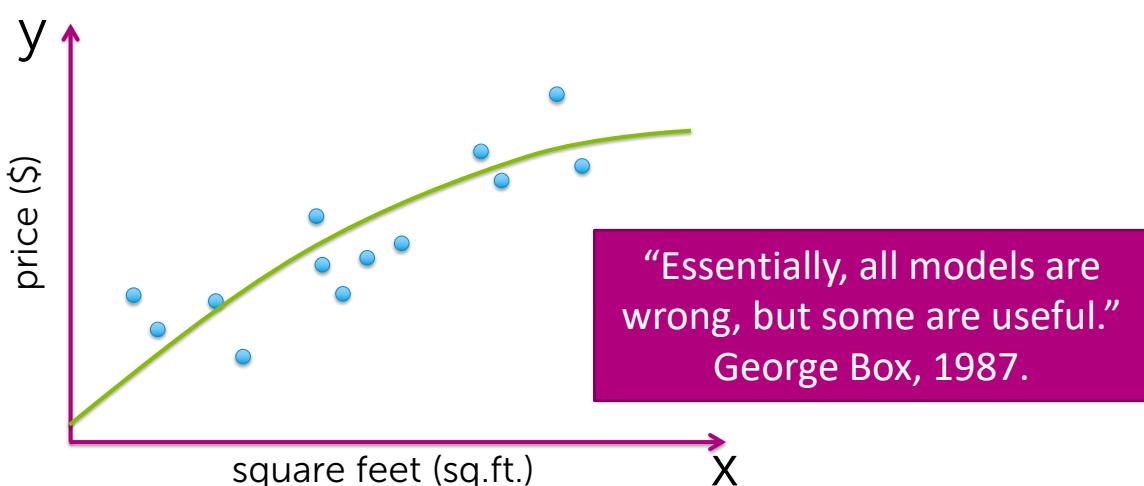
67

@2024 Emily Fox

CS 229: Machine Learning

67

Model – How we assume the world works



68

@2024 Emily Fox

CS 229: Machine Learning

68

Linear regression

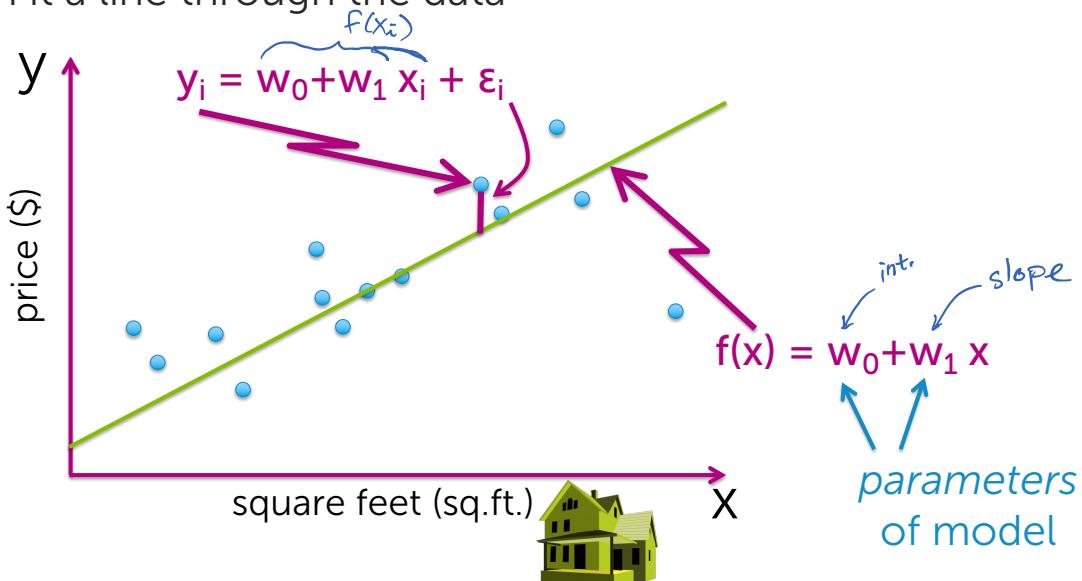
©2024 Emily Fox

CS 229: Machine Learning

69

Use a simple **linear** regression model

Fit a line through the data

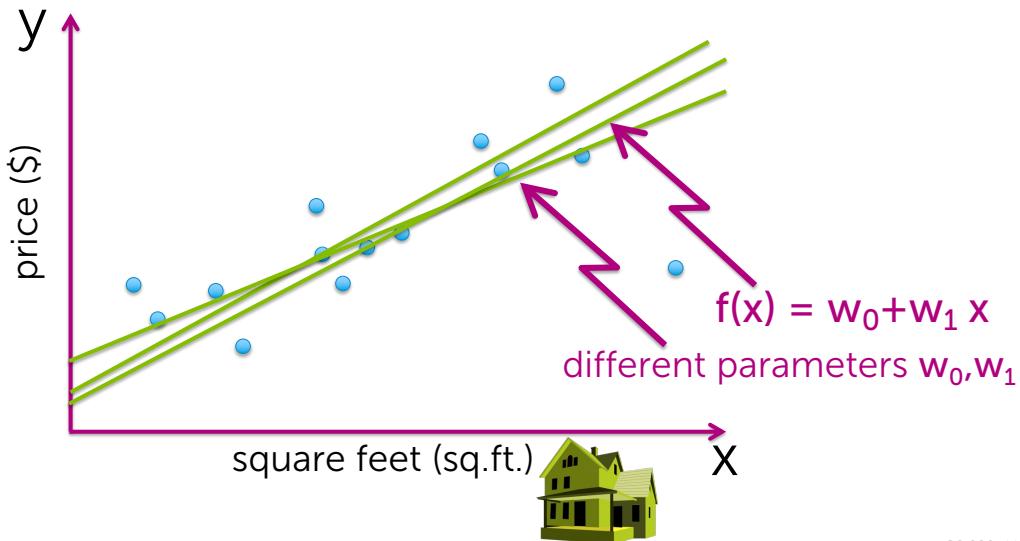


70

CS 229: Machine Learning

70

Which line?



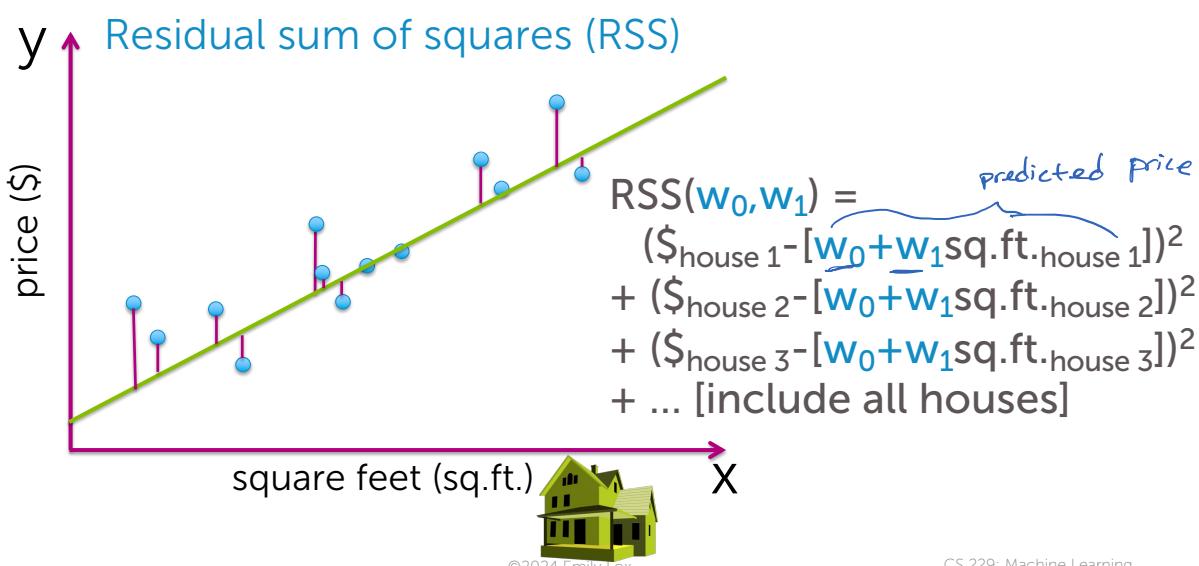
71

©2024 Emily Fox

CS 229: Machine Learning

71

"Cost" of using a given line (Loss function)



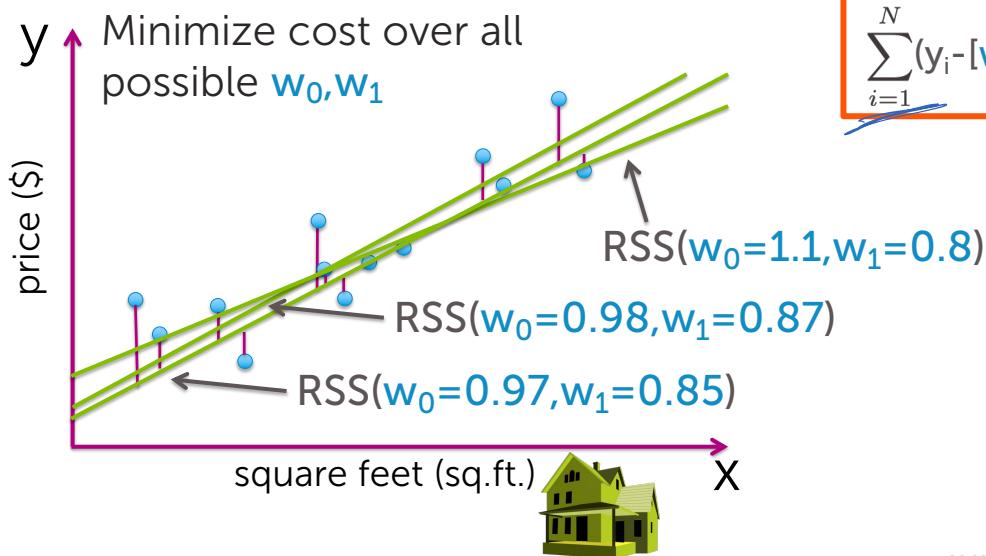
72

©2024 Emily Fox

CS 229: Machine Learning

72

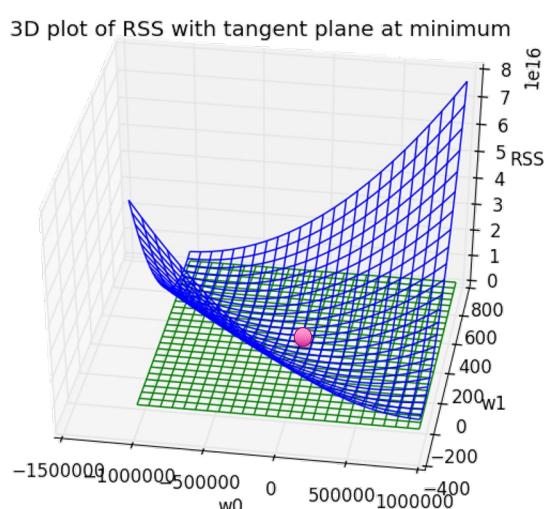
Find “best” line



73

73

Minimizing the cost



Minimize function over all possible w_0, w_1

$$\min_{w_0, w_1} \sum_{i=1}^N (y_i - [w_0 + w_1 x_i])^2$$

$RSS(w_0, w_1)$ is a function of 2 variables

74

74

CS 229: Machine Learning

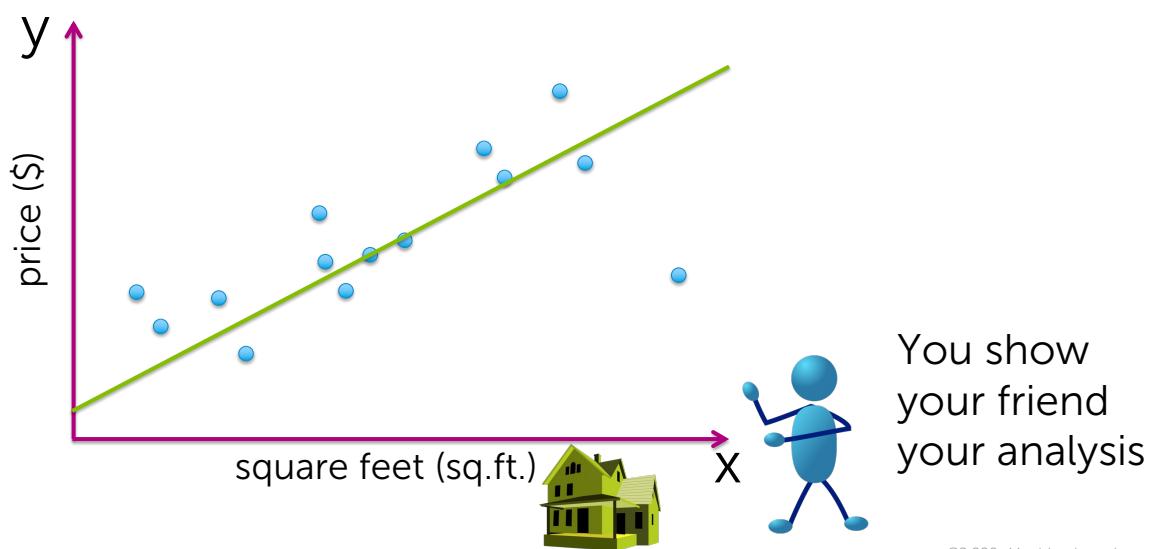
Adding higher order effects

©2024 Emily Fox

CS 229: Machine Learning

75

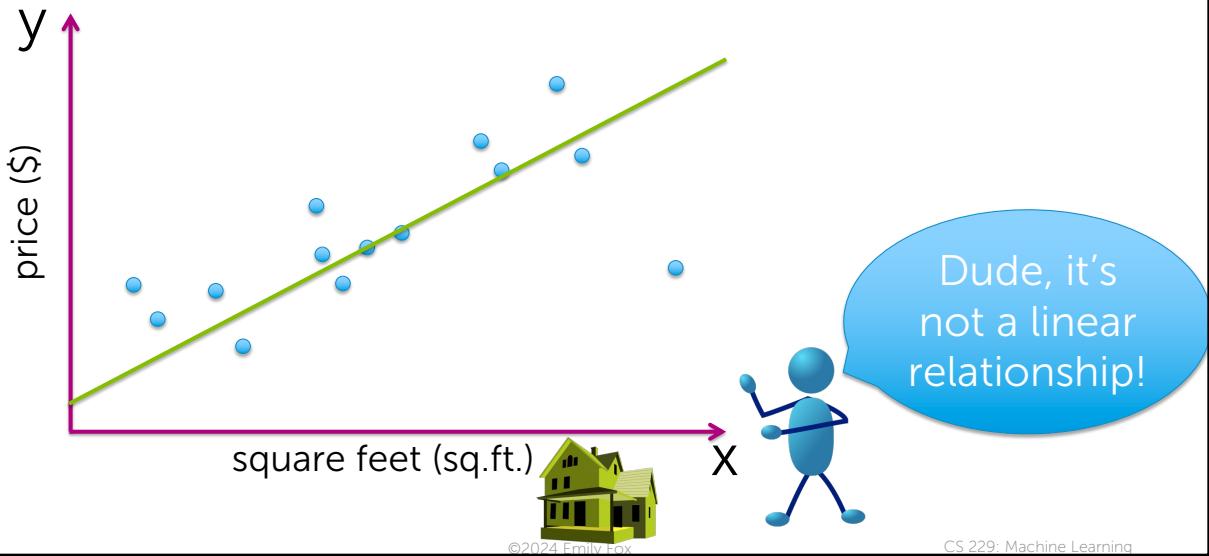
Fit data with a line or ... ?



76

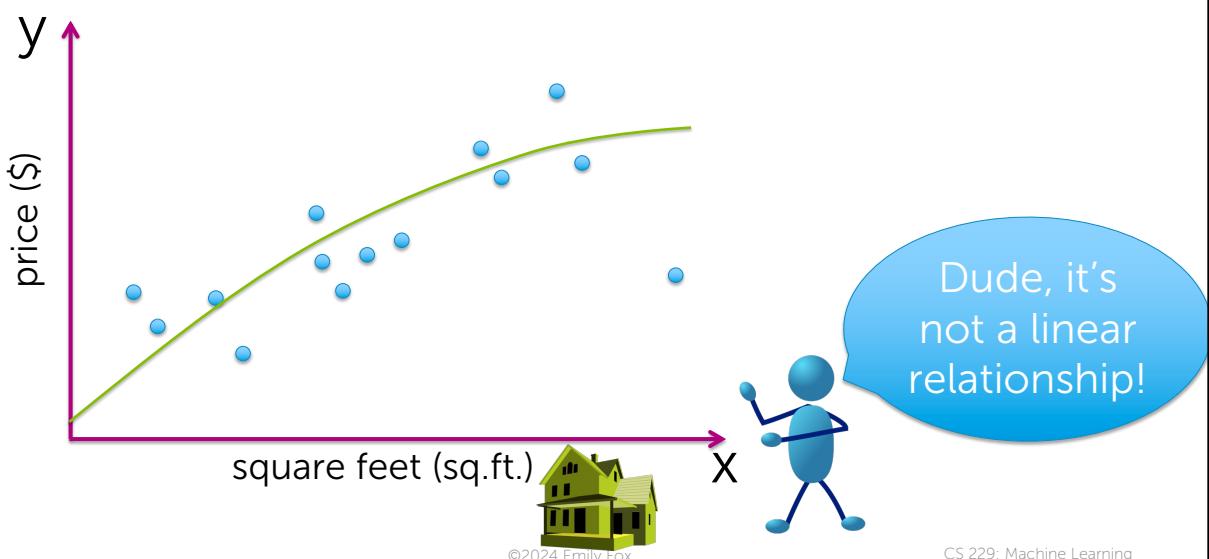
38

Fit data with a line or ... ?



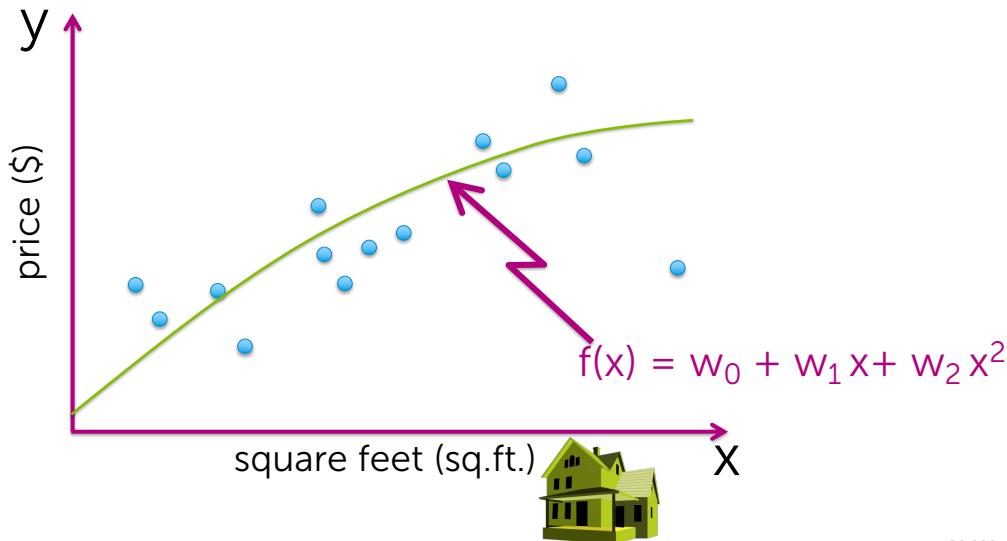
77

What about a quadratic function?



78

What about a quadratic function?

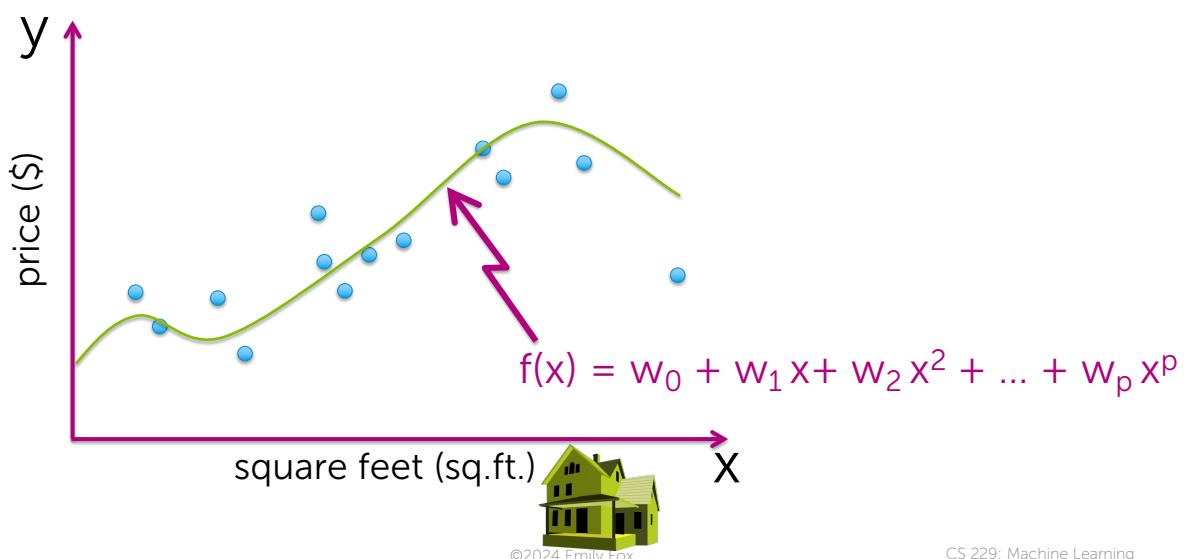


79

CS 229: Machine Learning

79

Even higher order polynomial



80

CS 229: Machine Learning

80

Polynomial regression

Model:

$$y_i = w_0 + w_1 x_i + w_2 x_i^2 + \dots + w_p x_i^p + \varepsilon_i$$

treat as different **features**

still "linear"
regression bc
linear function
of features

feature 1 = 1 (constant) parameter 1 = w₀

feature 2 = **x** parameter 2 = **w₁**

feature 3 = x^2 parameter 3 = w_2

3

feature $p+1 = \mathbf{x}^p$ parameter $p+1 = \mathbf{w}_p$

81

©2024 Emily Fox

CS 229: Machine Learning

Generic featurized linear regression

Model:

Model:

$$y_i = \mathbf{w}_0 h_0(x_i) + \mathbf{w}_1 h_1(x_i) + \dots + \mathbf{w}_D h_D(x_i) + \varepsilon_i$$

$$= \sum_{j=0}^D \mathbf{w}_j h_j(x_i) + \varepsilon_i$$

ith input (house)

jth feature

jth regression coefficient or weight

82

©2024 Emily Fox

CS 229: Machine Learning

82

Generic featurized linear regression

Model:

$$y_i = w_0 h_0(x_i) + w_1 h_1(x_i) + \dots + w_D h_D(x_i) + \varepsilon_i \\ = \sum_{j=0}^D w_j h_j(x_i) + \varepsilon_i$$

feature 1 = $h_0(x)$...often 1 (constant)

feature 2 = $h_1(x)$... e.g., x

feature 3 = $h_2(x)$... e.g., x^2 or $\sin(2\pi x/12)$ or $\log(x)$

...

feature $D+1 = h_D(x)$... e.g., x^p

83

©2024 Emily Fox

CS 229: Machine Learning

83

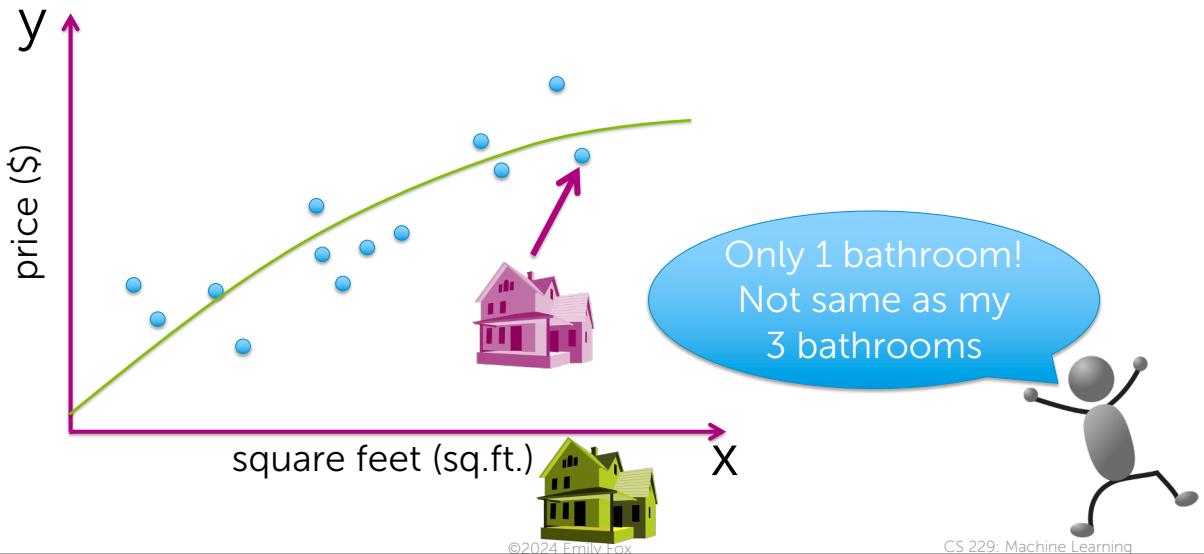
Adding other features

84

©2024 Emily Fox

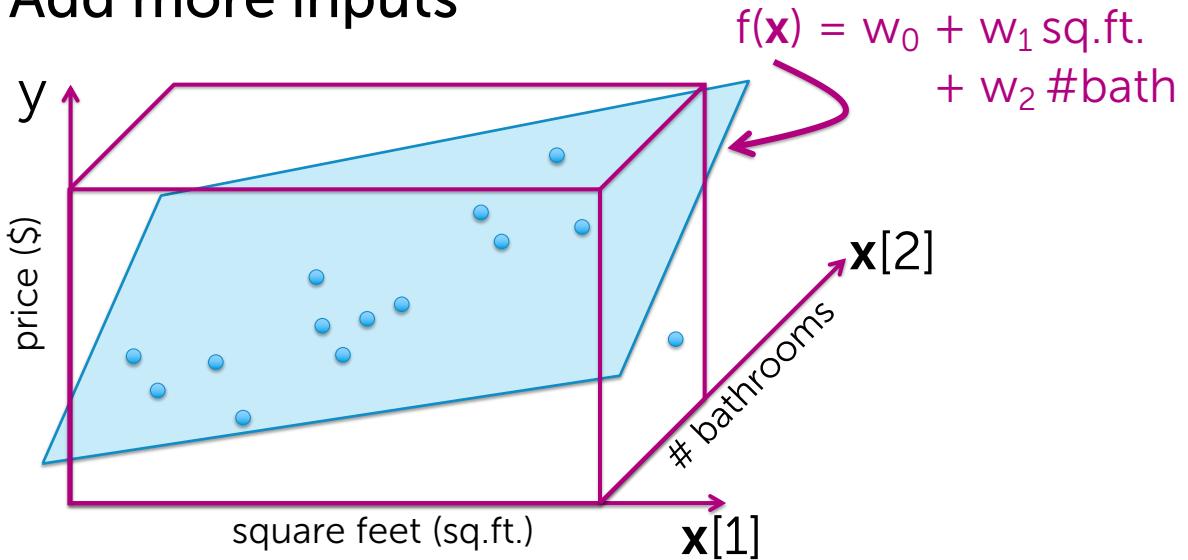
CS 229: Machine Learning

Predictions just based on house size



85

Add more inputs



86

86

Many possible inputs

- Square feet
- # bathrooms
- # bedrooms
- Lot size
- Year built
- Time since last remodel
- Zipcode
- ...

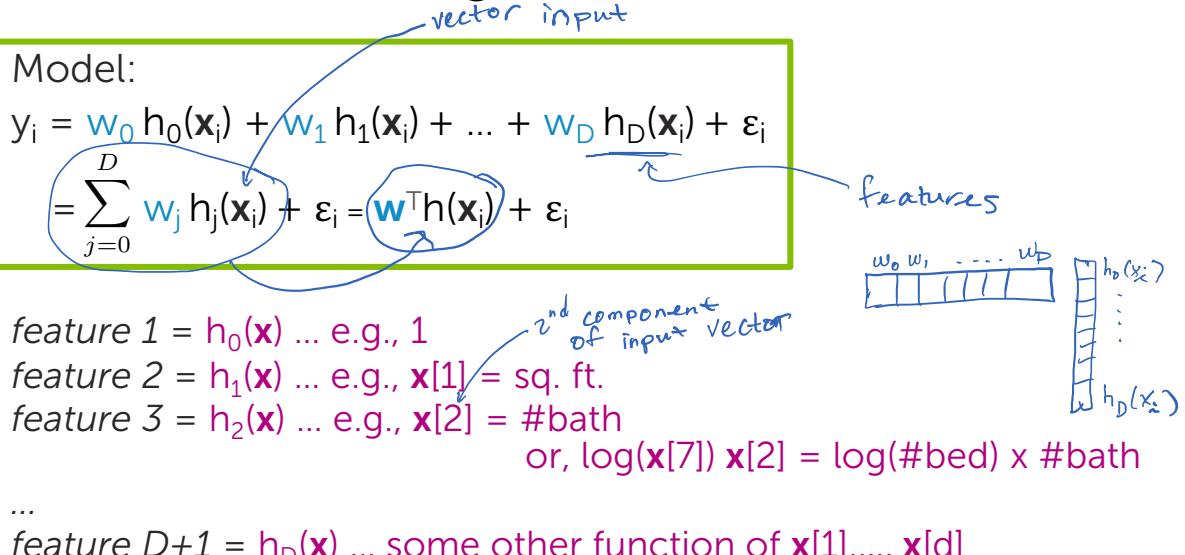
87

@2024 Emily Fox

CS 229: Machine Learning

87

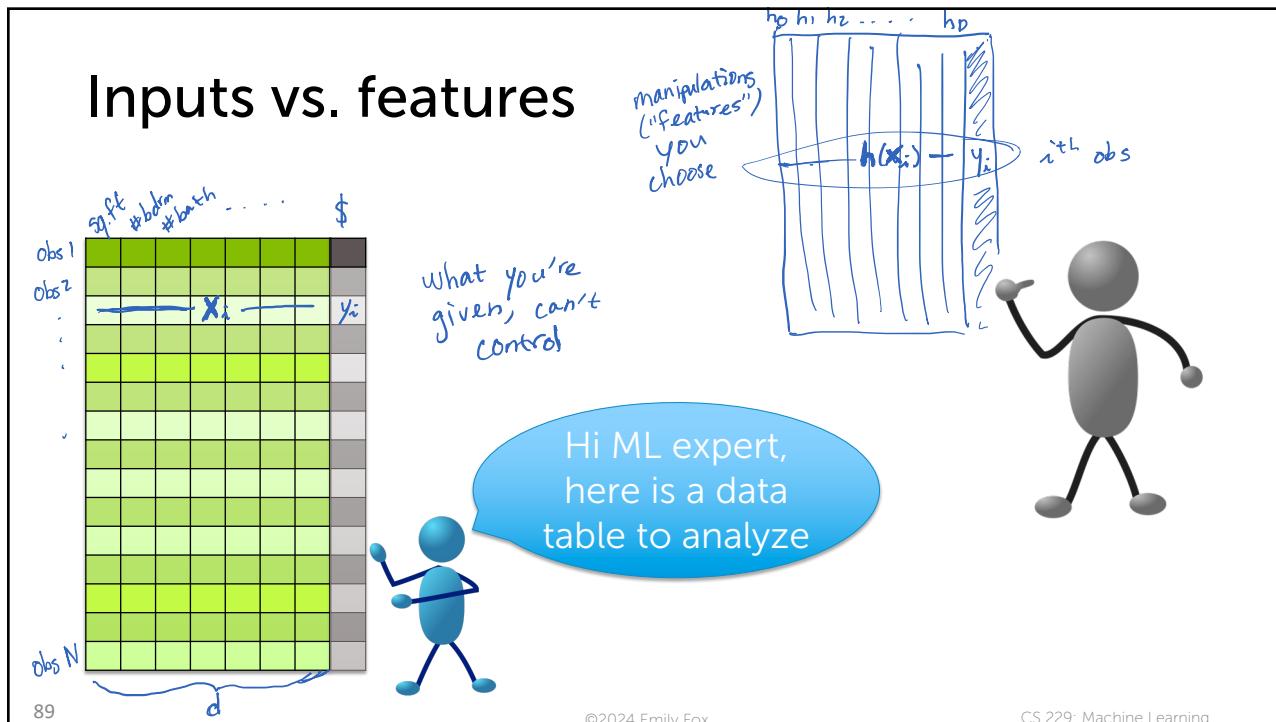
Generic linear regression model



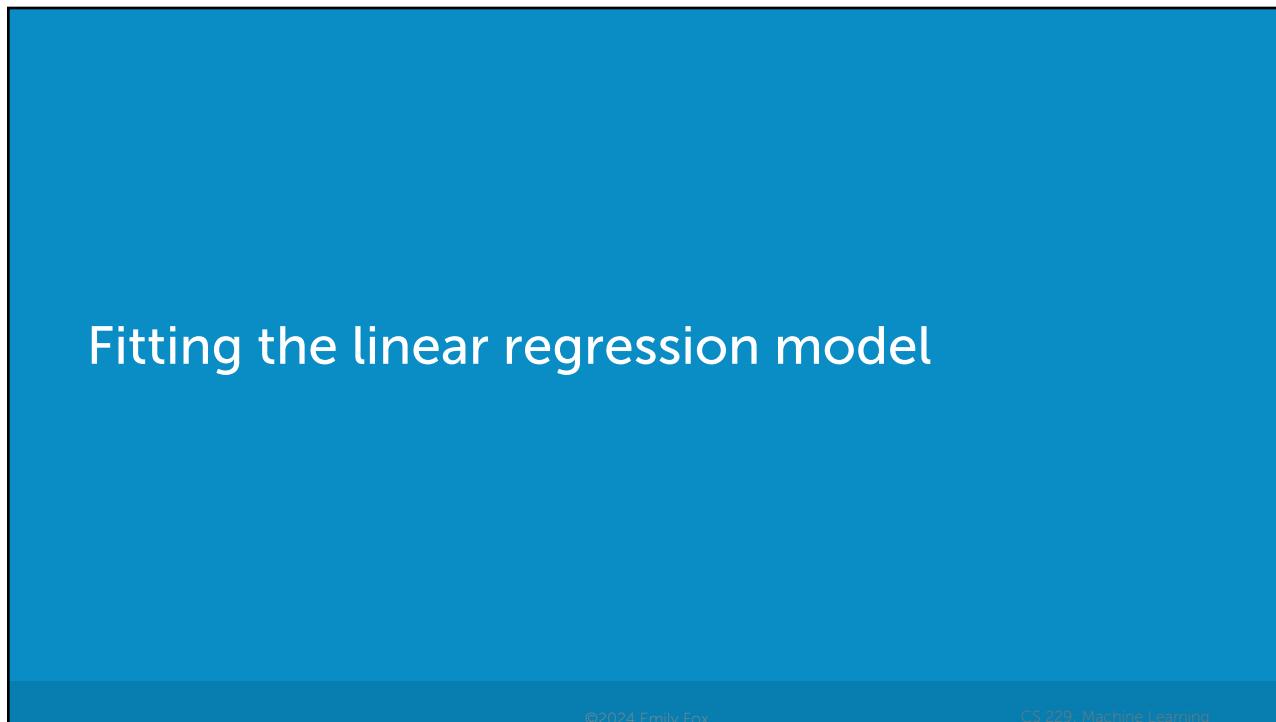
88

@2024 Emily Fox

CS 229: Machine Learning



89



90

Step 1: Rewrite the regression model

@2024 Emily Fox

CS 229: Machine Learning

91

Rewrite in matrix notation

For observation i

$$y_i = \sum_{j=0}^D w_j h_j(x_i) + \epsilon_i$$

$$y_i = \begin{bmatrix} w_0 & w_1 & \dots & w_D \end{bmatrix}^\top \begin{bmatrix} h_0(x_i) \\ h_1(x_i) \\ \vdots \\ h_D(x_i) \end{bmatrix} + \epsilon_i = \underbrace{\begin{bmatrix} h^\top(x_i) \end{bmatrix}}_{\text{matrix}} w + \epsilon_i$$

$$y_i = w^\top h(x_i) + \epsilon_i = \cancel{h^\top(x_i)} w + \epsilon_i$$

92

@2024 Emily Fox

CS 229: Machine Learning

Rewrite in matrix notation

For all observations together

$$\begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{matrix} = \begin{matrix} H \\ h(x_i) \end{matrix} + \begin{matrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{matrix} + \begin{matrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{matrix}$$

"design matrix"

$$\Rightarrow \boxed{y = Hw + \epsilon}$$

$$(y - Hw) = \epsilon$$

93

@2024 Emily Fox

CS 229: Machine Learning

93

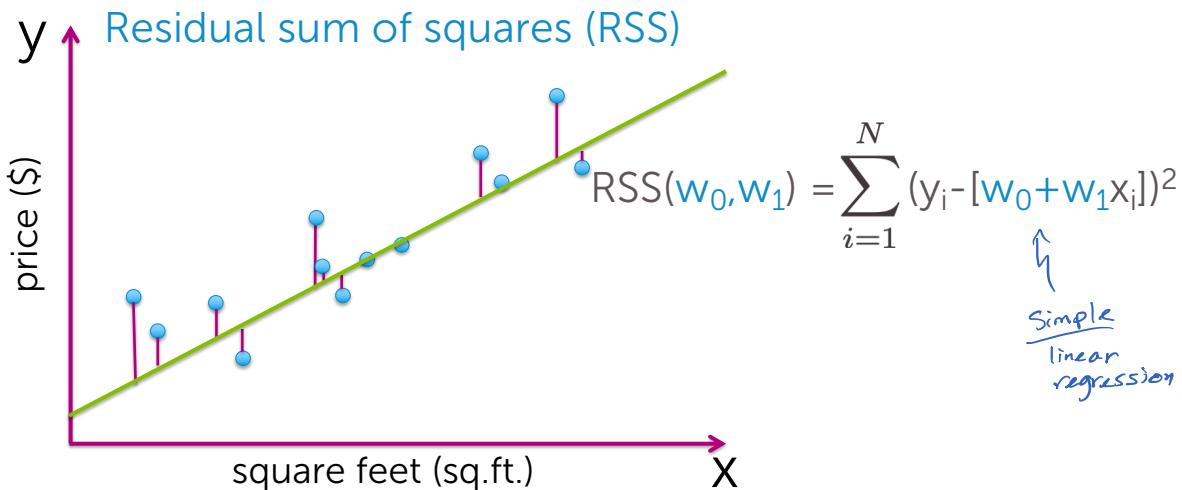
Step 2:
Compute the cost

@2024 Emily Fox

CS 229: Machine Learning

94

"Cost" of using a given line



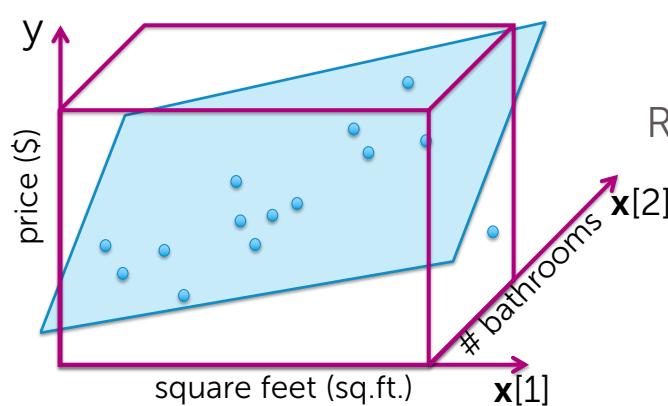
95

©2024 Emily Fox

CS 229: Machine Learning

95

RSS for multiple regression



$$\begin{aligned} \text{RSS}(\mathbf{w}) &= \sum_{i=1}^N (y_i - h^T(\mathbf{x}_i) \mathbf{w})^2 \\ &= (\mathbf{y} - \mathbf{H}\mathbf{w})^\top (\mathbf{y} - \mathbf{H}\mathbf{w}) \end{aligned}$$

actual obs
pred. value
 e_i

Show this at home

96

©2024 Emily Fox

CS 229: Machine Learning

96

Step 3:

Take the gradient

©2024 Emily Fox

CS 229: Machine Learning

97

Gradient of RSS

$$\nabla_{\mathbf{w}} \text{RSS}(\mathbf{w}) = \nabla[(\mathbf{y} - \mathbf{H}\mathbf{w})^\top (\mathbf{y} - \mathbf{H}\mathbf{w})]$$

$$= -2\mathbf{H}^\top (\mathbf{y} - \mathbf{H}\mathbf{w}) \quad \leftarrow \text{vector of dim } D+1 \quad \begin{matrix} \text{(length of } \\ \mathbf{w} \text{)} \end{matrix}$$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} \\ \frac{\partial f}{\partial x_{12}} \\ \vdots \\ \frac{\partial f}{\partial x_{1n}} \end{bmatrix}$$

vector of partials

Why? By analogy to 1D case:

$$\frac{d}{dw} (\mathbf{y} - \mathbf{h}\mathbf{w})(\mathbf{y} - \mathbf{h}\mathbf{w}) = \frac{d}{dw} (\mathbf{y} - \mathbf{h}\mathbf{w})^2 = 2(\mathbf{y} - \mathbf{h}\mathbf{w})(-\mathbf{h})$$

$$= -2\mathbf{h}^\top (\mathbf{y} - \mathbf{h}\mathbf{w}).$$

$\uparrow \uparrow$
scalars

98

©2024 Emily Fox

CS 229: Machine Learning

98

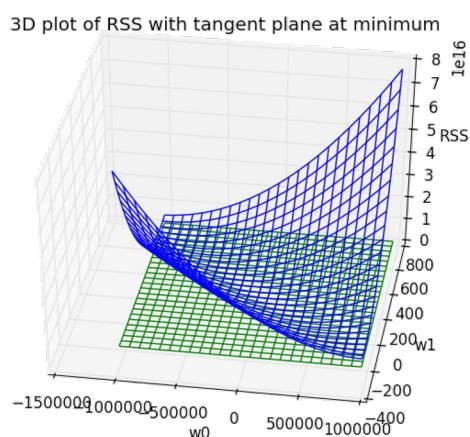
Step 4, Approach 1: Set the gradient = 0

©2024 Emily Fox

CS 229: Machine Learning

99

Closed-form solution



$$\nabla \text{RSS}(\mathbf{w}) = -2\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{w}) = 0$$

Solve for \mathbf{w} :

100

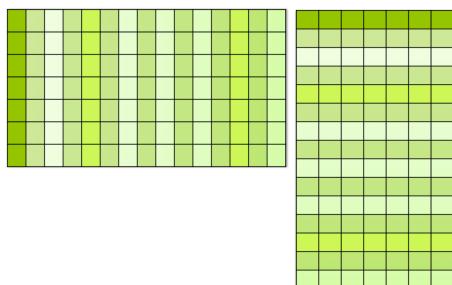
©2024 Emily Fox

CS 229: Machine Learning

100

Closed-form solution

$$\hat{\mathbf{w}} = \underbrace{(\mathbf{H}^T \mathbf{H})^{-1}}_{\downarrow} \mathbf{H}^T \mathbf{y}$$



Invertible if:

Complexity of inverse:

101

©2024 Emily Fox

CS 229: Machine Learning

101

Step 4, Approach 2: Gradient descent

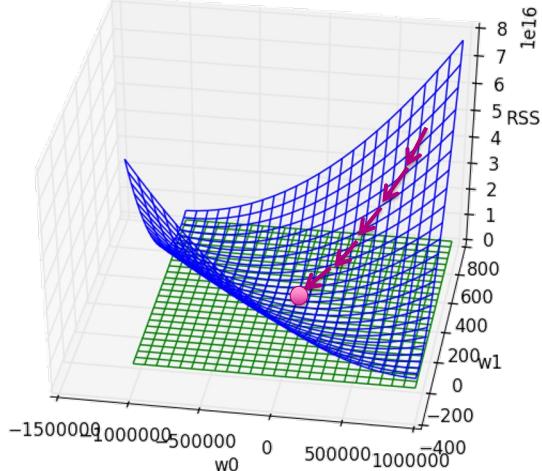
©2024 Emily Fox

CS 229: Machine Learning

102

Gradient descent

3D plot of RSS with tangent plane at minimum



©2024 Emily Fox

CS 229: Machine Learning

103

Algorithm:

while not converged

$$\mathbf{w}^{(t+1)} \leftarrow \mathbf{w}^{(t)} - \eta \nabla \text{RSS}(\mathbf{w}^{(t)})$$

\downarrow

$$-2\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{w})$$

$\hat{\mathbf{w}}$

103

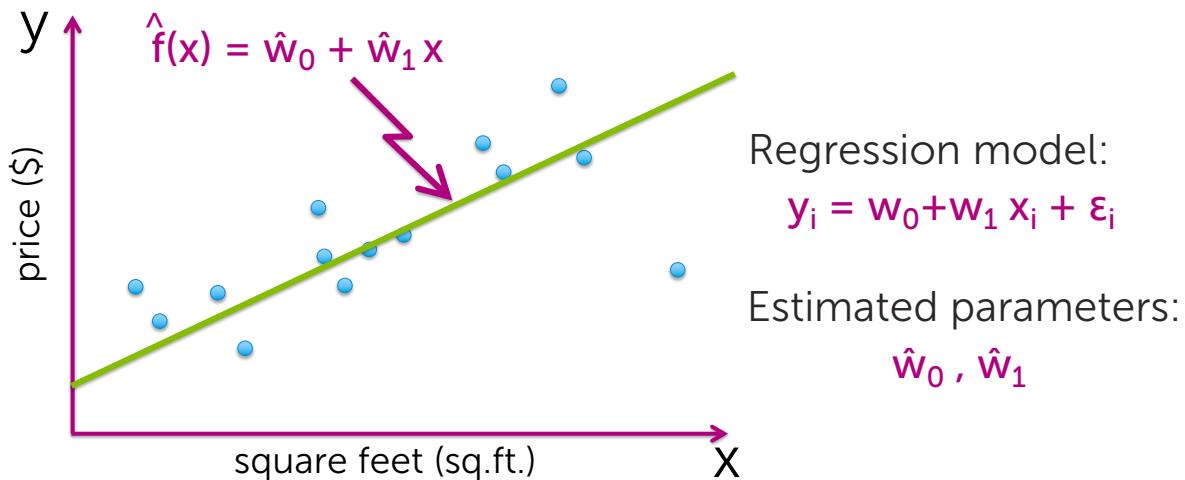
Using the fitted line

©2024 Emily Fox

CS 229: Machine Learning

104

Model vs. fitted line



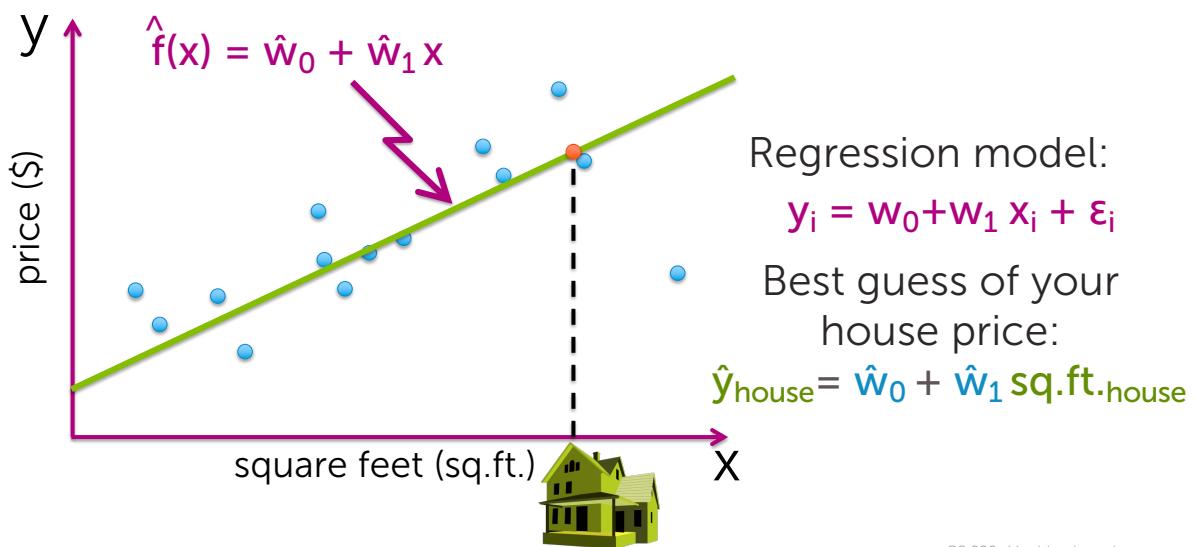
105

©2024 Emily Fox

CS 229: Machine Learning

105

Seller: Predicting your house price



106

©2024 Emily Fox

CS 229: Machine Learning

106

Summary of linear regression

©2024 Emily Fox

CS 229: Machine Learning

107

What you can do now...

- Describe the input (features) and output (real-valued predictions) of a regression model
- Write a regression model using multiple inputs or features thereof
- Cast both polynomial regression and regression with multiple inputs as regression with multiple features
- Calculate a cost metric (e.g., RSS)
- Estimate model parameters of a general multiple regression model to minimize RSS:
 - In closed form
 - Using an iterative gradient descent algorithm
- Exploit the estimated model to form predictions

108

©2024 Emily Fox

CS 229: Machine Learning

108

Appendix: Notation

Output: $y \leftarrow$ scalar

Inputs: $\mathbf{x} = (\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[d])$

\uparrow d-dim vector

Notational conventions:

$\mathbf{x}[j] = j^{\text{th}}$ input (scalar)

$h_j(\mathbf{x}) = j^{\text{th}}$ feature (scalar)

$\mathbf{x}_i =$ input of i^{th} data point (vector)

$\mathbf{x}_i[j] = j^{\text{th}}$ input of i^{th} data point (scalar)

109

@2024 Emily Fox

CS 229: Machine Learning

109

Appendix: More on notation

observations $(\mathbf{x}_i, y_i) : N$

inputs $\mathbf{x}[j] : d$

features $h_j(\mathbf{x}) : D$

110

@2024 Emily Fox

CS 229: Machine Learning

110