

L6: Linear Regression

Prof. Xun Jiao

Before class

- HW1 due by this Wed class
- HW2 will be posted on Wed

Outline

- Build a supervised learning model: a high-level perspective
- Linear Regression

Linear Regression Review regression supervised learning Machine learning classification unsupervised learning

Review on Classification Evaluation

		True/Actual		
		Positive (😭)	Negative	
Predicted	Positive (😭)	5	1	
	Negative	2	2	

Accuracy?

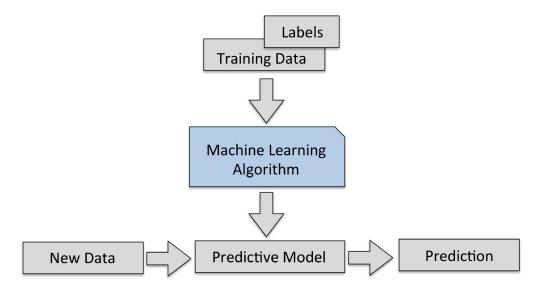
Precision?

Recall?

F1 score?

Now that we know how to evaluate a model, let us see how to build a model (supervised learning).

Supervised learning is the process of trying to infer from labeled data the underlying function that produced the labels associated with the data



- Predicting the future with supervised learning
- Classification vs. Regression

Given labeled training data of the form

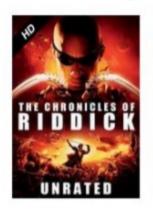
$$\{(\text{data}_1, \text{label}_1), \dots, (\text{data}_n, \text{label}_n)\}$$

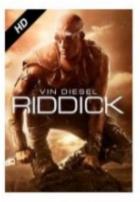
Infer the function

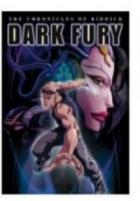
$$f(\text{data}) \stackrel{?}{\rightarrow} \text{labels}$$

Suppose we want to build a movie recommender

e.g. which of these films will I rate highest?

















Q: What are the labels?

A: ratings that others have given to each movie, and that I have given to other movies

333 of 115 people found the following review helpful
Excellent Sci 16. September 13. 2000
By Exic J. Pray (Sparate New York) - See all my reviews
This colors in force: Each Size (Sente (Sent
Rich Rich visit surgulatily one of the most conviolated filter of the early year. Although the esting of the filtri could seem nutries or a casual investigance transled and blokering on a housing stores in feetand with also managed. Among Taxini Taxini Taxini to an ordinary to receive the seem of the control country of the transless of the control country and the country of the transless of the country of the country of the transless of the country of the transless of the country of the transless of the country of the country of the transless of the country of the country of the transless of the country of the country of the country of the transless of the country of the cou
Visualized contribution that the failure inflammation interpretability in the failure inflammation interpretability in the failure inflammation in the failure inflammatio
The first does and counting great and rise insure than a fee, insurents of necessing sersion structure in the good measure. For access encount best does in a insurence, would all not the ment of you, by giving this first movine a charter.
may offer continuing find the med highly minima.
Was the review helpful to you? (No.) Committee (3)
27 of 40 people found the following review halful
(MARIN) Study research from the theater, September 5, 2000
Re Bazis Charles (Delve Dun, India) - San al my motions
Non-residue to France Park Mark Street Description Collisis
I star moving along a major to the contract of
Fig. 2 in the 19 to 1 and 19 t
as followed by Check it and by yourself and if you like it put the french around. Since it a character I had been to the follower and Three tripps. The world as seen through the eyes of the alone put me back in Princher's Allerd. While ultrasely the pint is fentage I like this monie take me away, in
much the same vary ve all let Star Warry (the frost three) move us and Starship Thospers moved us (appealed to our callective bissame failure). Have fun with it. I know I did.
ready order continues that the most height in minutes
Was the review highly to you? (No.) Sectional
33 of 39 people found the following review helpful
as or proposed states can entered processing the control of the co
to Kalley I fan al tra maiora
William Value of the Control of the
This is a derived good spirit florror flob. Class. It inn't ready access (more reaction inducing then review inducing and you do have to take it with a grain of salt, but in general, it is a very decent receive for the general. It was a finite little Alleron - same sort of feed, though not so humanous. (Wall, they start have a Husban, so you know).
The interest are given— not of like dragator as in facilities with a last of Barchier frequent in plant for the contract and a page of the facilities as contract as the delication and the facilities as contract as the facilities as contract as the facilities as the facilities as the facilities as for the fa
Fig. Riddith and Silving using excellent characters and both actors gave talented portrained. I use really exprised to learn that Cole Hauser use a Sinced 30 at the very least.
This is control characterist (Radio Michael and Europea, Claim House in a Solvin, and the Solvin House in a Solvin, and the Solvin House in a Solvin House in a Solvin House in a Solvin House in American in the Solvin House in American in the Solvin House in American in American in the Solvin House in American in the Solvin House in American
The print continues for the cost beginning to the cost of the cost
was the most region to your files. The Control of t
63 of 300 people found the following nevier-height
without Task, smart, organish filternaking, November 14, 2000
St. A Conference Code. Mark 1 (1994)
Hel has surely fragen over. That's the only you on explain how David Turby, under director of the sociacidity intercours (when skylving agic "Terminal Velocity," has made a movie this good. Dis not high and, but "Plantillatin" is a triumph within to general a suppressful, intelligent moview movie with supprisingly deep characters.
A demaged spacestry loaded with cargo and cyrevalesprog produces the loggest problem in the victors convex what is cargo and cyrevalesprog produces having beneath the planets surface. And then cores the loggest problem in the victors convex what is cargo and cyrevalesprog produces having beneath the planets surface. And then cores the total eclipse.
"Both State" is a Dissal dynaminated marking—bit to Dissals drawt in As at the measuring connect find factor P. Middle C. David gives a five ministration of the Dissals draw in a Dissals draw in the control of the Dissals draw in the control of the Dissals draw in the control of the Control
"Plich Black" is an emberrassment of riches for a cell fans characters who controlly surprise pox, creapy creatures Mrt mostly up to your imagnation, and a proposal device story that moves as a breakment, page. Reflect proporal entertainment— just be some you don't farm _pli_ the lights of before you creature.
Hilly other customers find this over high-furnishmer. Solid above. Solid above. Solid above. Solid above.
Was this review height to you? (Inc.) See Comment (1)
31 of 20 people found the following review helpful
A Silva St. A Silva SciFi Theriton, May 3, 2005
No A. Phellips - Lon all res. protects
8,8,100

Q: What is the data?

A: features about the movie and the users who evaluated it

Movie features: genre, actors, rating, length, etc.

Product Details

Genres	Science Fiction, Action, Horror			
Director	David Twohy			
Starring	Vin Diesel, Radha Mitchell			
Supporting actors	Cole Hauser, Keith David, Lewis Fitz-Gerald, Claudia Black, Rhiana G Angela Moore, Peter Chiang, Ken Twohy			
Studio	NBC Universal			
MPAA rating	R (Restricted)			
Captions and subtitles	English Details 🔻			
Rental rights	24 hour viewing period. Details ▼			
Purchase rights	Stream instantly and download to 2 locations Details 🔻			
Format	Amazon Instant Video (streaming online video and digital download)			

User features: age, gender, A. Phillips location, etc.

Reviewer ranking: #17,230,554

90% helpful

votes received on reviews (151 of 167)

ABOUT ME

Enjoy the reviews...

ACTIVITIES

Reviews (16)

Public Wish List (2)

Listmania Lists (2)

Tagged Items (1)

What is data? What is label?

No.	Model	Luxury?	Year	MPG	Horsepower	Price
1	Acura MDX	Yes	2017	20	290	\$50,000
2	Honda Accord	No	2017	25	190	\$25,000
3	Honda Civic	No	2012	23	160	\$10,000
4	Honda Civic	No	2016	24	170	\$18,000
5	Nissan Altima	No	2016	30	180	\$25,000
6	Acura MDX	Yes	2015	18	280	\$38,000
7	Lexus RX350	Yes	2015	21	270	\$40,000
8	Toyota Prius	No	2014	45	120	\$28,000
9	Toyota Prius	No	2013	40	120	\$24,000

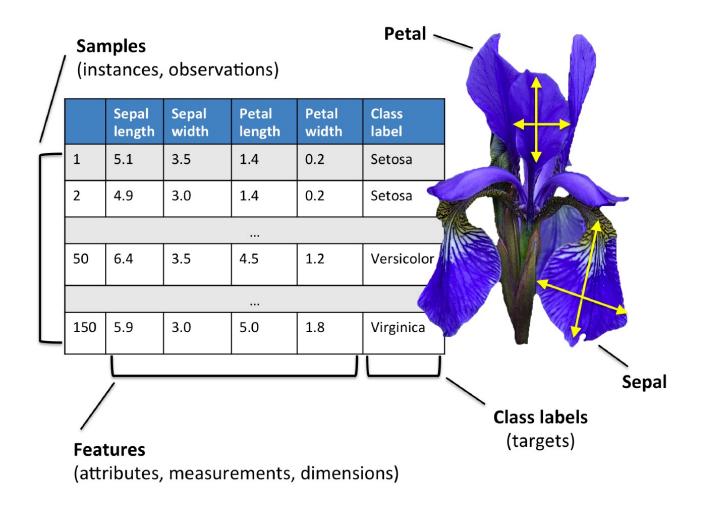
Movie recommendation:

$$f(\text{data}) \stackrel{?}{\rightarrow} \text{labels}$$

=

 $f(\text{user features}, \text{movie features}) \stackrel{?}{\rightarrow} \text{star rating}$

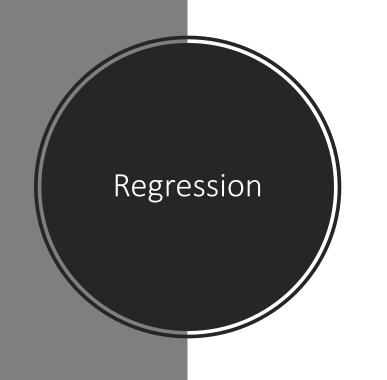
A classic example: iris dataset



Basic Terminology

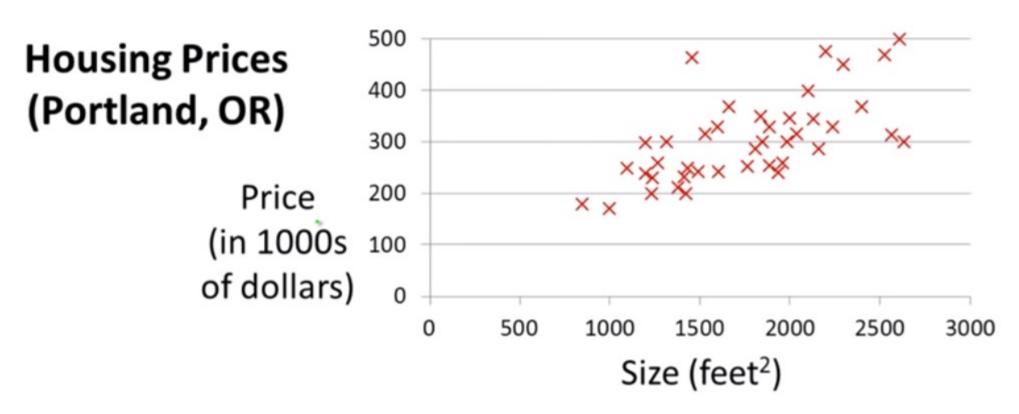
- Measurements of 150 iris flowers (150 samples / 4 features)
- From 3 different species (Setosa, Versicolor, Virginica)
- Rows are samples and columns are features
- 150×4 matrix $\mathbf{X} \in \mathbb{R}^{150 \times 4}$:

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_4^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{(150)} & x_2^{(150)} & x_3^{(150)} & \dots & x_4^{(150)} \end{bmatrix}$$



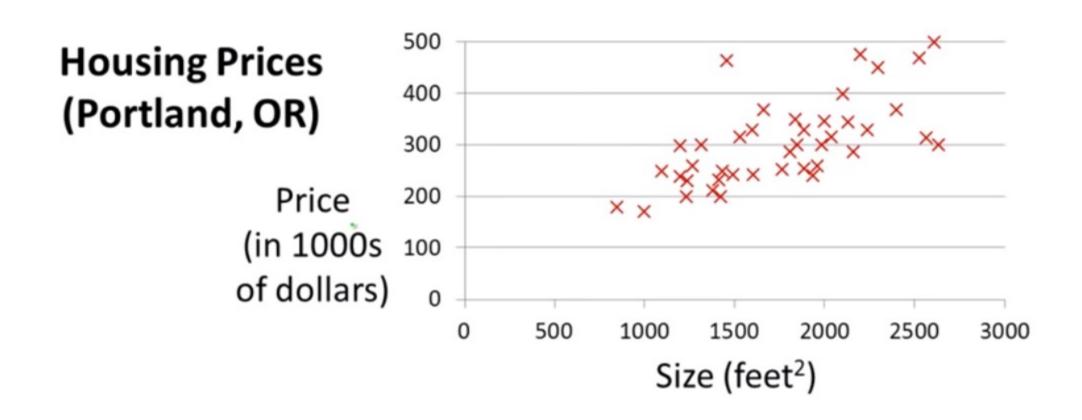
Regression is one of the simplest supervised learning approaches to learn relationships between input variables (features) and output variables (predictions)

A regression example



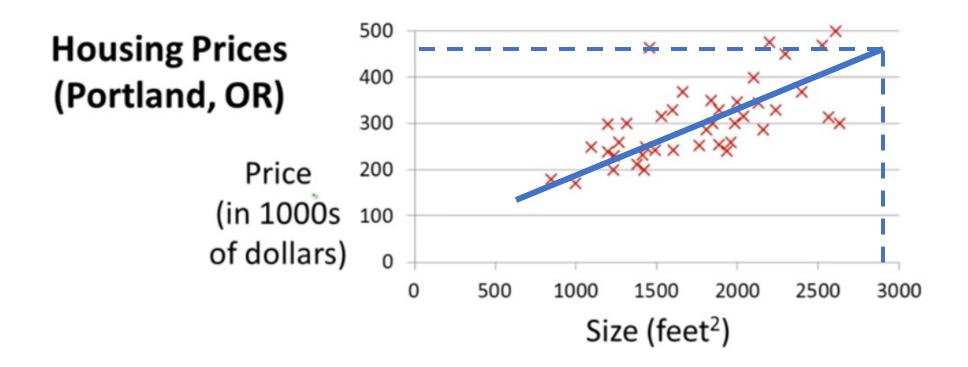
Suppose I want to build a predictive model that can predict "Price" based on "Size"
What is my input? Output?

So, how can we find such model/function?



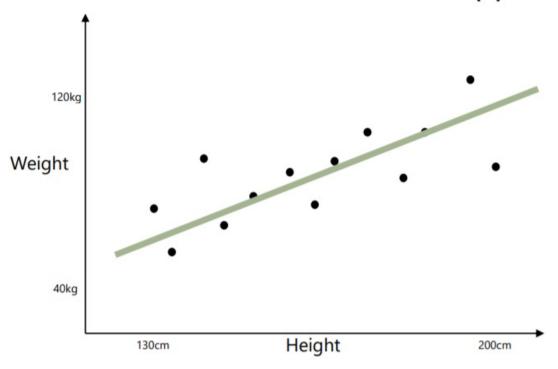
The easiest assumption

 The easiest assumption would be a linear relationship (mapping) from data to label.



Motivation: height vs. weight

Q: Can we find a line that (approximately) fits the data?



Motivation: height vs. weight

Q: Can we find a line that (approximately) fits the data?

- If we can find such a line, we can use it to make predictions (i.e., estimate a person's weight given their height)
 - How do we formulate the problem of finding a line?
 - If no line will fit the data exactly, how to approximate?
 - What is the "best" line?

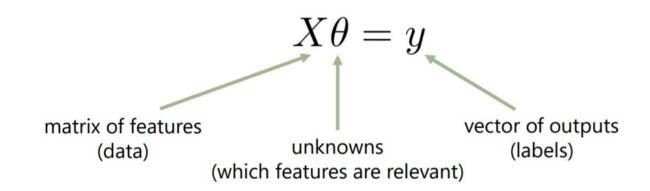
Goal of regression

- Given a set of training data, find the mapping function f that can
 - Fit the training data
 - Map from X -> y
 - Input data → label

To answer these questions, we need to look into the general mathematical form of linear relationship.

Linear Regression

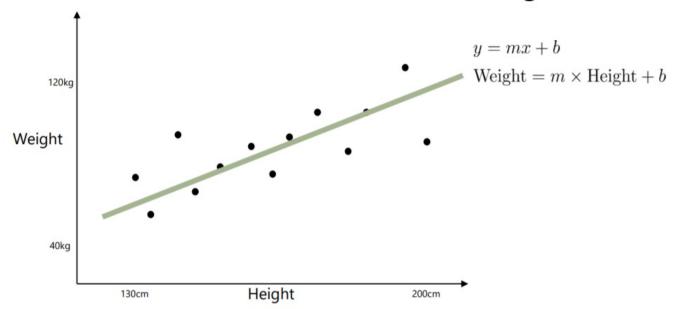
Linear regression assumes a predictor of the form



(or Ax = b if you prefer)

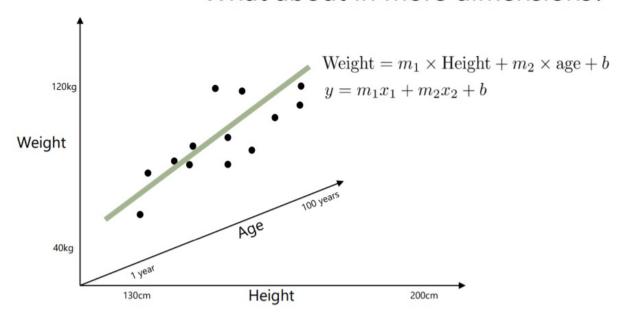
Recap: equation for a line

What is the formula describing the line?



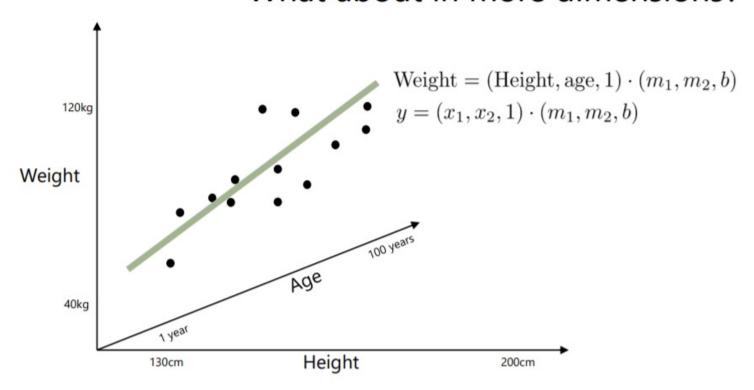
Recap: equation for a line

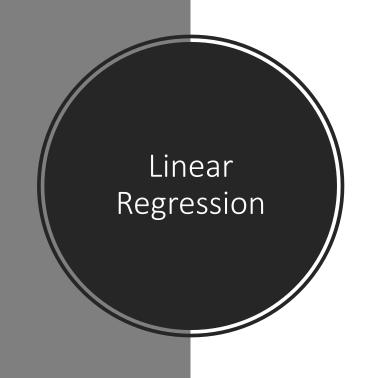
What about in more dimensions?



Recap: equation for a line as an inner product

What about in more dimensions?





Linear regression assumes a predictor of the form

$$X\theta = y$$

Q: Solve for theta