Intro to Linear Regression

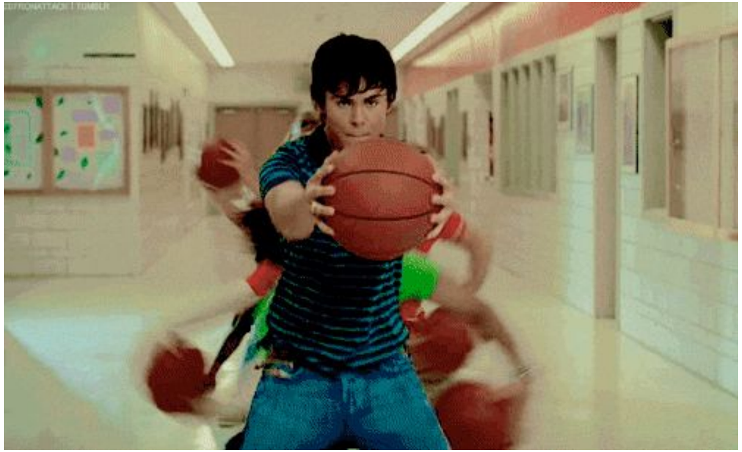
Session 4

Slides Link:
https://teachla.uclaacm.com/classes/ml

ACM AI + ACM TeachLA

What is your favorite weather?

Recap



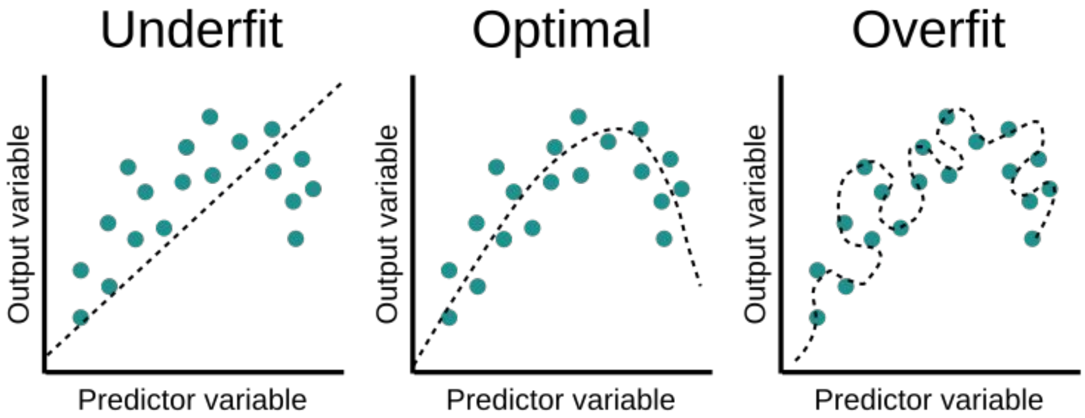
Supervised Machine Learning

- A Machine Learning *model* is a big function
- A Supervised Learning model uses labeled output data, hence the name “supervised”
- Supervised Learning **maps the relationship of:**

$$\text{input} \rightarrow \text{labeled output}$$

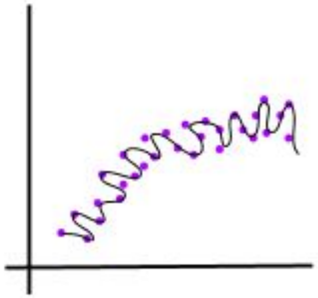
Training

- ML Models "*train*": finding the relation between input → output
- Train on A LOT of diverse data to avoid **overfitting/underfitting**

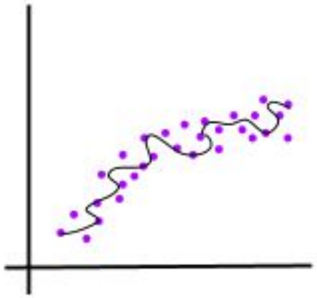


Review

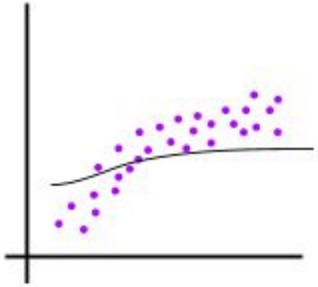
Look at this photoGRAPH!



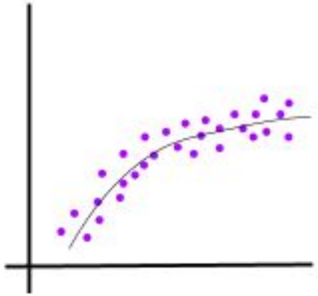
i)________________



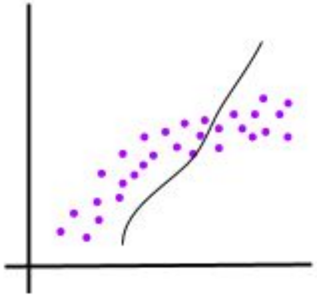
ii)________________



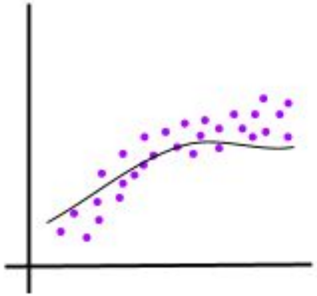
iii)________________



iv)________________

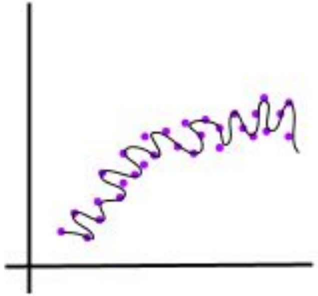


v)________________

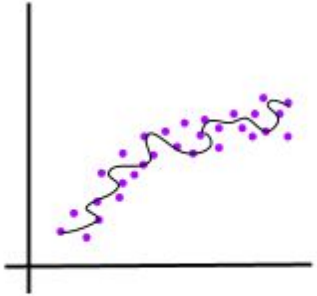


vi)________________

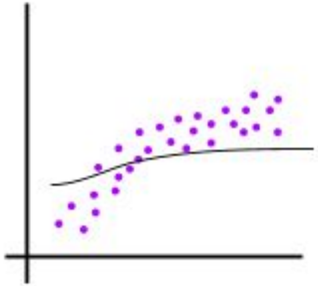
Look at this photoGRAPH (answers)!



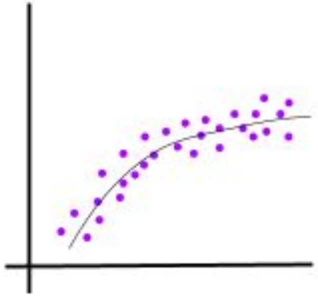
i) Overfitting



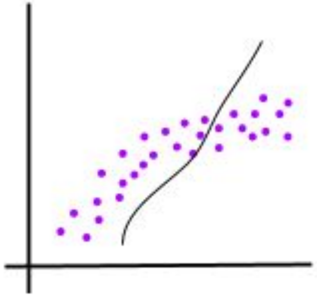
ii) Overfitting



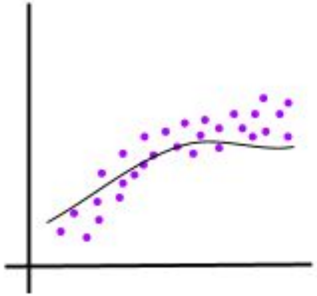
iii) Underfitting



iv) Optimal



v) Underfitting



vi) Optimal

and some ~ questions ~

- How can we avoid overfitting?
- How can we avoid underfitting?
- What is the difference between the training dataset and the test dataset?

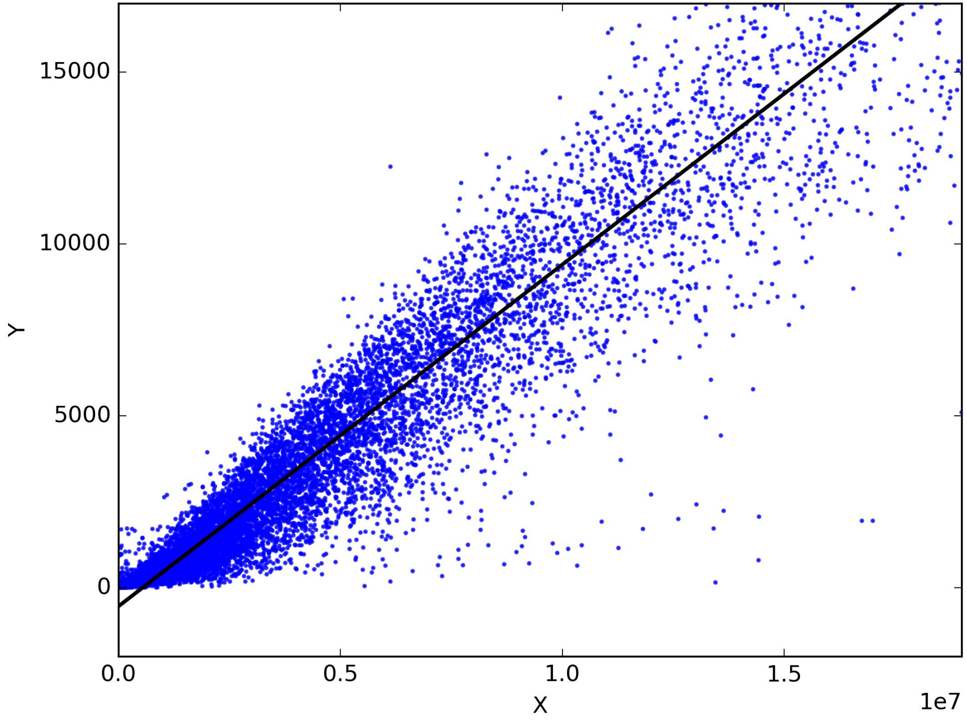
and some ~ questions ~ (answers)

- How can we avoid overfitting?
  - Use a diverse dataset + more data
- How can we avoid underfitting?
  - Use more data
- What is the difference between the training dataset and the test dataset?
  - We use the training dataset to find a relation between the input and output. We use the testing dataset to evaluate how accurate that relation is after the model has been trained

Questions?

Linear Regression

What is linear regression?



What is linear regression?

Regression refers to the relationship between two or more variables.

- Data from the real world **rarely** carries a relationship among its variables that can be **modeled perfectly** by a function such as a linear function, a quadratic, an exponential, etc.
- Let's look at an example!

What is linear regression?

Here is some data on the heights and weights of NFL players from 2014.
Assume that a taller player usually weighs more than a shorter player.

NFL player	Position	Height (inches)	Weight (pounds)
Jahleel Addae	Defense	71	195
Tim Benford	Offense	71	196
Victor Butler	Defense	74	231
Hebron Fangupo	Offense	73	330
Anthony Fasano	Offense	76	255
Brian Hartline	Offense	74	180
Jason Hatcher	Defense	79	285
Cullen Jenkins	Defense	75	292
Darrin Reaves	Offense	70	210
Scott Simonson	Offense	77	249
Aldon Smith	Defense	77	255
Isaiah Trufant	Defense	68	170

Table 2: Heights and weights of a simple random sample of NFL players from the 2014 season. Data from pro-football-reference.com.

What is linear regression?

Do you think we can use a linear function that relates **all** the heights (x) to weights (y)?

A. Yes

B. No

NFL player	Position	Height (inches)	Weight (pounds)
Jahleel Addae	Defense	71	195
Tim Benford	Offense	71	196
Victor Butler	Defense	74	231
Hebron Fangupo	Offense	73	330
Anthony Fasano	Offense	76	255
Brian Hartline	Offense	74	180
Jason Hatcher	Defense	79	285
Cullen Jenkins	Defense	75	292
Darrin Reaves	Offense	70	210
Scott Simonson	Offense	77	249
Aldon Smith	Defense	77	255
Isaiah Trufant	Defense	68	170

Table 2: Heights and weights of a simple random sample of NFL players from the 2014 season. Data from pro-football-reference.com.

What is linear regression?

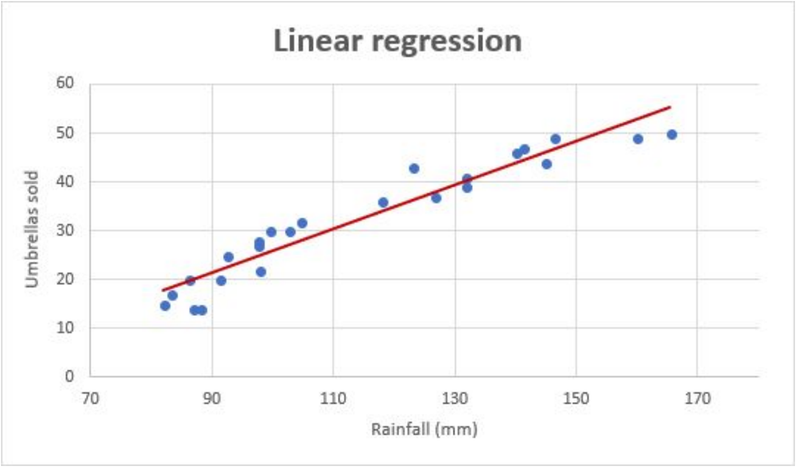
You can tell by looking at the first two players that we can't have a function that contains all these points because Jahleel Addae and Tim Benford are both 71 inches tall but have different weights.

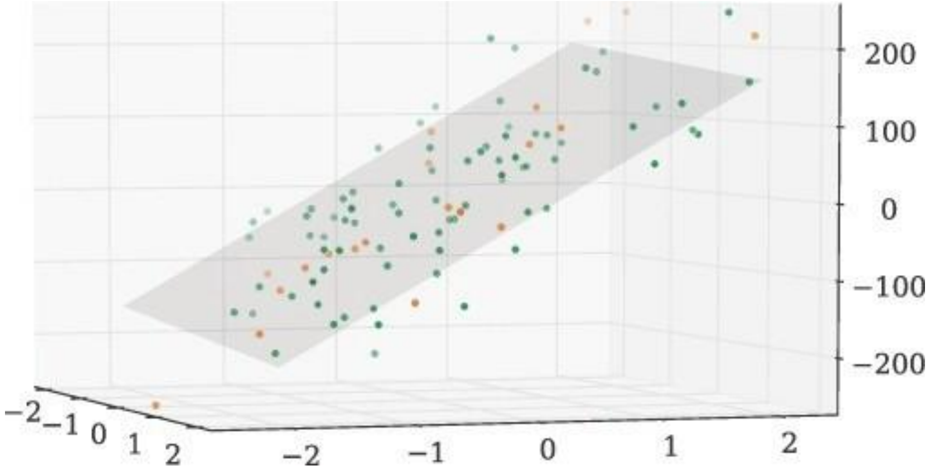
NFL player	Position	Height (inches)	Weight (pounds)
Jahleel Addae	Defense	71	195
Tim Benford	Offense	71	196
Victor Butler	Defense	74	231
Hebron Fangupo	Offense	73	330
Anthony Fasano	Offense	76	255
Brian Hartline	Offense	74	180
Jason Hatcher	Defense	79	285
Cullen Jenkins	Defense	75	292
Darrin Reaves	Offense	70	210
Scott Simonson	Offense	77	249
Aldon Smith	Defense	77	255
Isaiah Trufant	Defense	68	170

Table 2: Heights and weights of a simple random sample of NFL players from the 2014 season. Data from pro-football-reference.com.

What is linear regression?

- We want to find a way to take in all our data and create one function that best fits it.
- In case of **linear** regression, that function is a line or a plane!



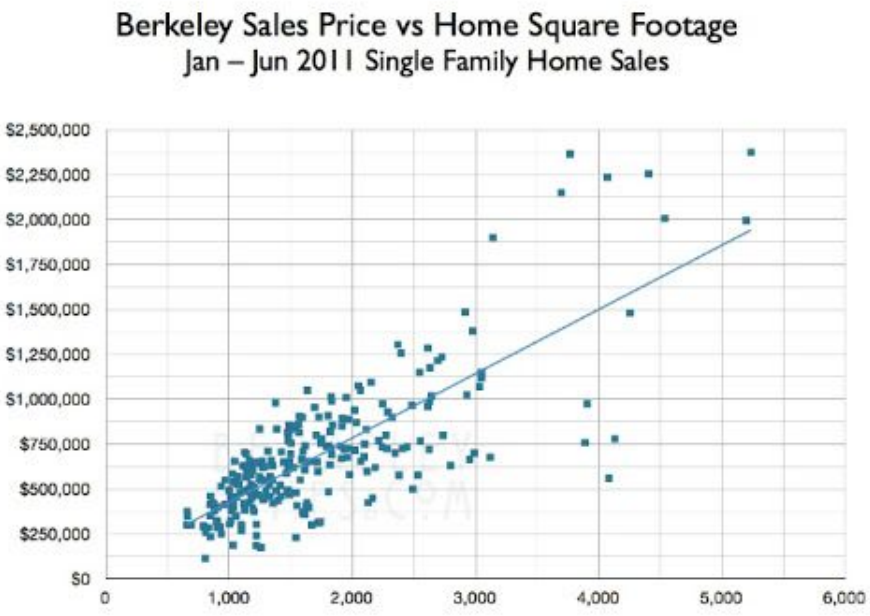


What is linear regression?

The Goal: Create a line or plane that models the universal relationship between several inputs and one output.

This is a big idea in DL: **universal approximation!**

- aka Line of Best Fit!



Classic example: Using a line to model the <u>relationship</u> between square footage and property price

Food for thought

1. Why use <u>linear regression</u>?

2. Why do we want <u>diversity</u> in our training data?

Linear Regression, formally

Let's write the training dataset like this:

$[(x_0, y_0),$ ⇦ Start at 0, not 1

$(x_1, y_1),$

$\ldots,$

$(x_i, y_i),$ ⇦ When talking about <u>any</u> given point, we say it is the i-th training example

$\ldots,$

$(x_m, y_m)]$ ⇦ m total training examples

What do <u>x and y</u> represent in this graph on the right?

Linear Regression, formally

What do x and y represent in this graph on the right?

Answer:

- x = home square footage
- y = sales price (of the home)

We say that home square footage is an input feature.

What other features can we use?

Linear Regression, formally

Some possible features:

- Size of the home (ft^2)
- Age of the home (yr)
- Distance to nearest shopping center (mi)

$$\begin{pmatrix} 2000\ \text{ft}^2 \\ 20\ \text{yr} \\ 2\ \text{mi} \end{pmatrix}$$

← This is a vector!

Quick Aside: Vectors

$2000\ \text{ft}^2$

- A scalar

$$\begin{pmatrix} 2000\ \text{ft}^2 \\ 20\ \text{yr} \\ 2\ \text{mi} \end{pmatrix}$$

- 3 features
- 3-dimensional

$$\begin{bmatrix} 2000\ \text{ft}^2 & 20\ \text{yr} & 2\ \text{mi} \end{bmatrix}$$

- Can also be written like this!

More generally… $\left[\ x_0, x_1, \ldots, x_i, \ldots, x_n\ \right]$

Linear Regression, formally

Let's put together what we have so far...

1. For some <u>input vector</u> x_i, the corresponding output is the <u>scalar</u> y_i
2. *Example*: say we pick (x_2, y_2)

$$x_2 = \begin{pmatrix} 2000 \\ 20 \\ 2 \end{pmatrix}, \quad y_2 = 1{,}500{,}000$$

Linear Regression, formally

Say that some training dataset looks like this:

$$x_0 = \begin{bmatrix} 12 & 3 & 56 & 700 \end{bmatrix}, \quad y_0 = 5$$

$$x_1 = \begin{bmatrix} 9 & 6 & 52 & 703 \end{bmatrix}, \quad y_1 = 7$$

$$x_2 = \begin{bmatrix} 11 & 5 & 51 & 698 \end{bmatrix}, \quad y_2 = 4$$

- How many training examples (m)?
  - 3
- How many features (n)?
  - 4
  - Input vectors are "4-dimensional"

Linear Regression, formally

$$x_0 = \begin{bmatrix} 12 & 3 & 56 & 700 \end{bmatrix}, \quad y_0 = 5$$

$$x_1 = \begin{bmatrix} 9 & 6 & 52 & 703 \end{bmatrix}, \quad y_1 = 7$$

$$x_2 = \begin{bmatrix} 11 & 5 & 51 & 698 \end{bmatrix}, \quad y_2 = 4$$

This is *gross*!
YUCK!

<ins>Can we write the training dataset more cleanly?</ins>

Yes! We can <ins>stack</ins> the input vectors to form a matrix.

Linear Regression, formally

$$x_0 = \begin{bmatrix} 12 & 3 & 56 & 700 \end{bmatrix}, \quad y_0 = 5$$
$$x_1 = \begin{bmatrix} 9 & 6 & 52 & 703 \end{bmatrix}, \quad y_1 = 7$$
$$x_2 = \begin{bmatrix} 11 & 5 & 51 & 698 \end{bmatrix}, \quad y_2 = 4$$

$$Y = \begin{bmatrix} 5 \\ 7 \\ 4 \end{bmatrix} \quad m = 3$$

$$X = \begin{bmatrix} 12 & 3 & 56 & 700 \\ 9 & 6 & 52 & 703 \\ 11 & 5 & 51 & 698 \end{bmatrix} \quad m = 3$$

$$n = 4$$

Stacked vectors
= a matrix

The dimensions
(m, n)
stay the same!

But why do we
write it like this?

Linear Regression, formally

The Linear Regression problem can be written as $\boldsymbol{Y} = W\boldsymbol{X} + B$.
We will learn more about how to calculate **Y** in future lessons!

1. What are the dimensions of matrix **X**?

2. What are the dimensions of vector **Y**?

Our Goal

Linear Regression: $Y = WX + B$

- Find <u>best possible</u> **W** and **B**
- Such that our regression model outputs $Y_{predict}$ very close to $Y_{reality}$

Quick recap: putting it all together

Input data → **Preprocessing** → **Training** → **Testing** → **Interpretation**

Input data: Training and testing data––the more diverse and more data we have, the better

Preprocessing: Scaling data, removing outliers, etc.

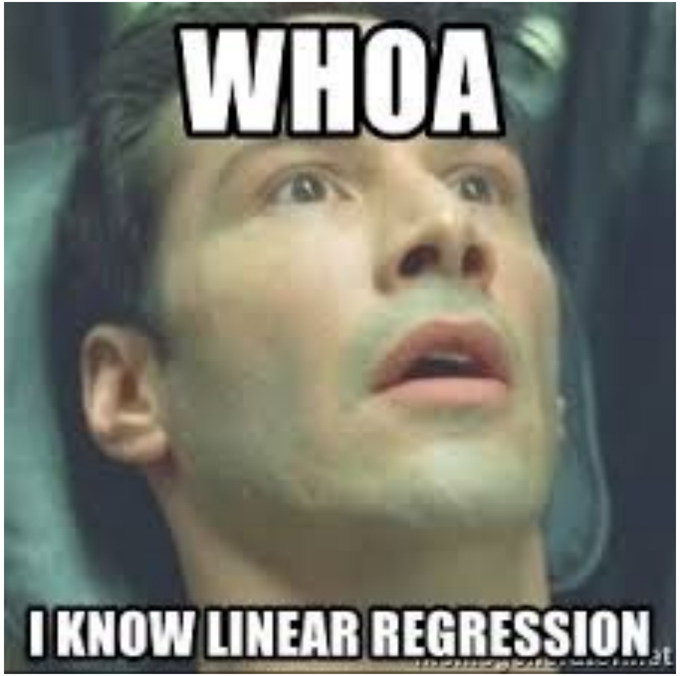
Training: Choosing the **function** and representing the data in that form––training data is used to develop model

Testing: Evaluating the model using testing data

Interpretation: Choosing a new model, introducing more variables, etc.

Function/model: maps input data (training examples) to outputs (predictions)
Example: y = mx + b

That's it for now, we'll look into the math behind linear regression soon!



Kahoot!

ACM AI

Thanks!

Fill out our Feedback Form:

Check out our Github
Read our Blog
ucla.acm.ai@gmail.com