Deep Dive into
Machine Learning

Karthikeyan NG

December 2017

What Society thinks I do

What Friends think I do

What Parents think I do

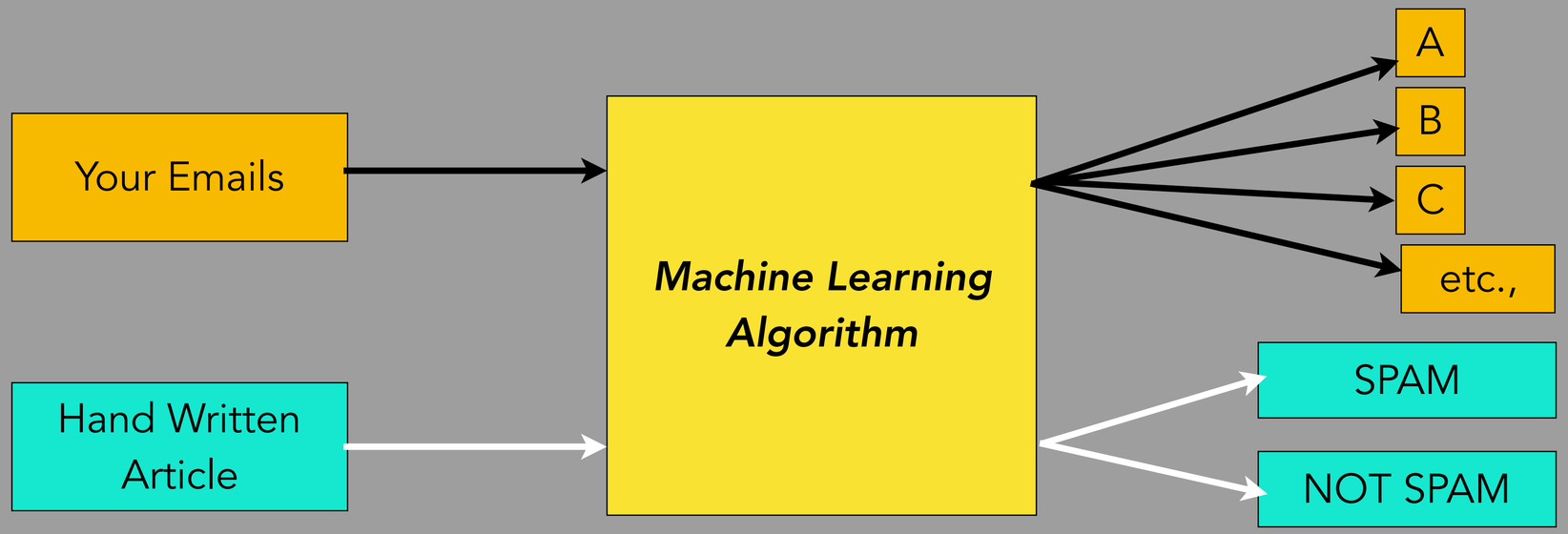
What Programmers think I do

What I think I do

```
➜  ~ python3
Python 3.6.3 (default, Oct  4 2017, 06:09:
[GCC 4.2.1 Compatible Apple LLVM 9.0.0 (cl
Type "help", "copyright", "credits" or "li
>>> import tensorflow as tf
/usr/local/Cellar/python3/3.6.3/Frameworks
portlib/_bootstrap.py:219: RuntimeWarning:
python.framework.fast_tensor_util' does no
  return f(*args, **kwds)
>>>
```

What I actually do

“Machine Learning is a concept where set of generic algorithms analyses your data and provides you interesting data without writing any specific code for your problem”



Machine Learning

Types of ML

- Supervised Learning
- Unsupervised Learning

Supervised Learning

You have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data

Unsupervised Learning

You only have input data (X) and no corresponding output variables.

The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

Steps in ML

1. Data collection
2. Data preparation
3. Choosing a model
4. Training
5. Evaluation
6. Parameter tuning
7. Prediction

Supervised Learning

No of Rooms	Square Feet	Location	Price
3	2300	New Delhi	₹ 7500000
4	2700	Bangalore	₹ 8000000
2	1800	Mumbai	₹ 6000000
1	750	Chennai	₹ 3500000
2	1300	Hyderabad	₹ 4000000
3	1650	Kolkata	₹ 4200000

Real estate training data

Supervised Learning

No of Rooms	Square Feet	Location	Price
3	1900	New Delhi	₹ ?

Can we predict this data now?

Unsupervised Learning

No of Rooms	Square Feet	Location
3	2300	New Delhi
4	2700	Bangalore
2	1800	Mumbai
1	750	Chennai
2	1300	Hyderabad
3	1650	Kolkata

Real estate training data

Let’s put this into code!

```
def estimate_house_sales_price(num_of_rooms, sqft, location):
  price = 0
  # In my area, the average house costs 2000 per sqft
  price_per_sqft = 2000
  if location == "bangalore":
    # but some areas cost a bit more
    price_per_sqft = 5000
  elif location == "chennai":
    # and some areas cost less
    price_per_sqft = 4000
  # start with a base price estimate based on how big the place is
  price = price_per_sqft * sqft
  # now adjust our estimate based on the number of bedrooms
  if num_of_rooms == 0:
    # Studio apartments are cheap
    price = price  - 1000000
  else:
    # places with more bedrooms are usually
    # more valuable
    price = price + (num_of_rooms * 500000)
 return price
```

Let’s put this into code!

```
def estimate_house_sales_price(num_of_rooms, sqft, location):
  price = < Computer, please DO MAGIC HERE >
 return price
```

This is what I want!

Let’s put this into code!

```
def estimate_house_sales_price(num_of_rooms, sqft, location):
 price = 0
 #by doing this
 price += num_of_rooms * .634873962437
 #and this
 price += sqft * 1875.4324324
 #maybe a this too
 price += location * 3.343345768
 #and finally, just a little extra salt for good measure
 price += 191.23434328
 return price
```

What are all these numbers?

Let’s find those numbers step-by-step

```
def estimate_house_sales_price(num_of_rooms, sqft, location):
 price = 0
 #by doing this
 price += num_of_rooms * 1.0
 #and this
 price += sqft * 1.0
 #maybe a this too
 price += location * 1.0
 #and finally, just a little extra salt for good measure
 price += 1.0
 return price
```

Step 1: Let’s start with weight 1.0

Let’s find those numbers step-by-step

No of	Square Feet	Location	Price	New Price
3	2300	New Delhi	₹ 7500000	₹ 9305670
4	2700	Bangalore	₹ 8000000	₹ 4900500
2	1800	Mumbai	₹ 6000000	₹ 5200230
1	750	Chennai	₹ 3500000	₹ 5507500
2	1300	Hyderabad	₹ 4000000	₹ 8000300
3	1650	Kolkata	₹ 4200000	₹ 3200750

Step 2: Calculate price with the new algorithm

Let’s find those numbers step-by-step

- Find the different between actual price and new price
- Find difference for all the houses listed
- The total amount is your total deviation
- Now work towards bring the difference value to zero
- Try each and every combination of weights
- When we have the difference closer to zero, boom! We have solved the problem!

Step 3: Do step 2 till we reach optimum value

Can we optimise the logic now?

$$\text{Cost} = \frac{\sum_{i=1}^{500} \left(\text{MyGuess}(i) - \text{RealAnswer}(i)\right)^2}{500 \cdot 2}$$

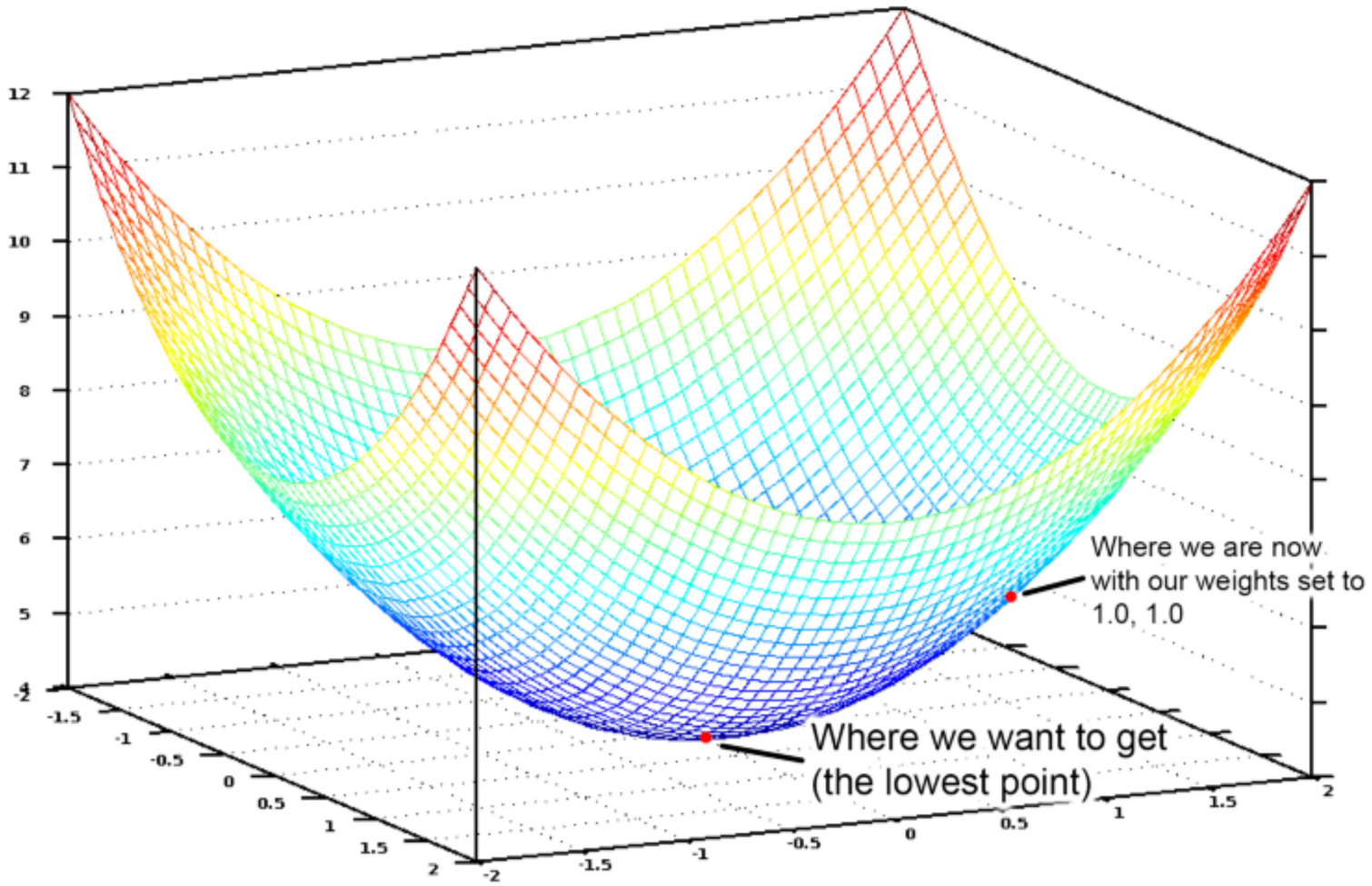
Cost function

Can we optimise the logic now?

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

θ is what represents your current weights. J(θ) means the ‘cost for your current weights’.

Can we optimise the logic now?



The graph of our cost function looks like a bowl. The vertical axis represents the cost.

Thanks!