

The background is a gradient from deep red at the top to dark blue at the bottom, speckled with white dots resembling stars. Overlaid on this are several faint, white circular patterns. Some are solid lines, while others are dashed. Some circles have arrows indicating a clockwise direction. One large circle on the left has a scale with numbers from 140 to 260 in increments of 10. Other smaller circles are scattered across the frame, some with partial segments highlighted.

# IN DEPTH MACHINE LEARNING

WITH PYTHON

“ Definition: Field of study that gives computer the ability to learn without being explicitly programmed

”  
A program is said to learn from experience  $E$ , with respect to task  $T$  and some performance measure  $P$  if its performance measure on  $T$  as measured by  $P$  improves with experience  $E$ .

# DIVISIONS :

## 1. Supervised

- a. Regression Problem
- b. Classification Problem

## 2. Learning Theory

## 3. Unsupervised Learning

## 4. Reinforcement Learning

# AIM OF THE CLASS (SUPERVISED LEARNING- REGRESSION)

- 1.Linear Regression(Theory)
- 2.Optimising Techniques :
  - a. Batch Gradient Decent(Theory)
  - b. Stochastic Gradient Decent(Theory)
  - c. Normal Equation(Theory)
- 3. Locally Weighted Regression(Theory)
- 4. Newtons Method(Theory)
- 5. Simple Linear Regression model fit without libraries(Python Code)
- 6. Linear Regression With some feature engineering(Python Code)
- 7. Using different optimizing Algorithms for the above problems
- 8. Introduction to Scikit-learn and using it to fit the above data.(Python code)



# SUPERVISED LEARNING (PREDICTIVE LEARNING)



# TERMINOLOGY:

- 1. Training Set:  $\mathcal{D} = \{(x_i, y_i)\}_{i=1 \text{ to } n}$
- 2. Fitting Parameter :  $\theta$
- 3. Hypothesis function :  $h_{\theta}(x)$
- 4. Cost Function :  $J(\theta)$
- 5. Learning Rate :  $\alpha$

# LINEAR REGRESSION

- Lets start with an example.....

# MULTI VARIABLE REGRESSION

Let's extend previous example.....



# OPTIMIZING TECHNIQUES

- 1. Gradient Descent(Batch Gradient Descent)
  - 2. Stochastic Gradient Descent
  - 3. Normal Equation
- 4. Newton's Method(We will discuss tomorrow)

# NORMAL EQUATION:

- WHY?
- HOW?

# BATCH GRADIENT DESCENT

Why?

How?

Why not?



# STOCHASTIC GRADIENT DESCENT

Why?

How?

Why not?





# NORMAL EQUATION

- Let's:
  - Vectorize all the parameters, and sample data to remove loops
    - Introduce  $\nabla$
- We will use the following properties of trace of matrix to find  $\Theta$ :
  - $\text{tr}(AB) = \text{tr}(BA)$
  - $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$ 
    - $\nabla_{\text{A}} \text{tr}(AB) = B^*$
    - $\text{tr}(A) = \text{tr}(A^*)$
  - $\nabla_{\text{A}} \text{tr}(ABA^*C) = CAB + C^*AB^*$

NORMAL EQUATION:

$$(X^*X)\theta = (X^*Y)$$

Take Pseudo inverse to get  $\theta$ .

# A LITTLE FEATURE ENGINEERING....

- 1. What is underfitting?
- 2. What is overfitting?
- 3. How can we increase the efficiency of a program?

# LOCALLY WEIGHED REGRESSION (LOESS/LOWESS ALGORITHM)

- Why?
- When ?
- How?





# CODE TIME

