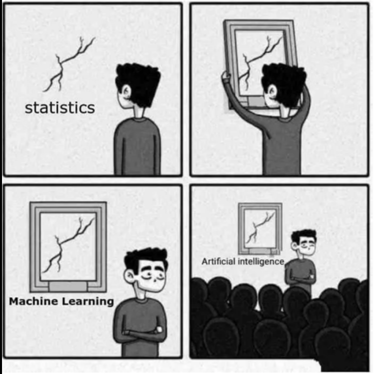
Week 03- Session 01 - Regression



Kelvin C. - PassDowns/Walkthroughs III

Table Of Content

Supervised Learning Overview

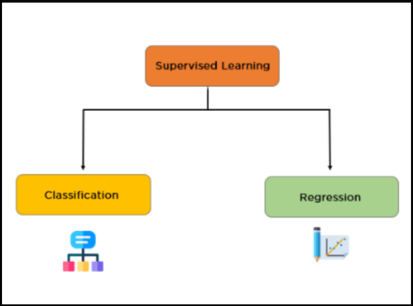


Figure: Illustration of Classification.

- **Supervised learning:** Learning a function that maps inputs to outputs based on labeled data.
- Two main types:
  - **Classification:** Predict discrete labels.
  - **Regression:** Predict continuous values.

Regression: Problem Setup

Regression on predicting house price

- Goal: Predict a continuous output y given input features $\mathbf{x}$.
- Examples:
  - Predict house prices based on features like size, location.
  - Predict temperature based on time of day.

Univariate Linear Regression

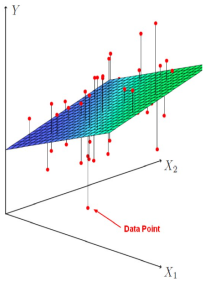
- One input feature x, output y.
- Hypothesis function:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

- θ_0: intercept, θ_1: slope.

Multivariate Linear Regression

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11



Regression on predicting house price

- Multiple input features $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
- Hypothesis function:

$$h_\theta(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

- Vectorized form:

$$h_\theta(\mathbf{x}) = \theta^T \mathbf{x}$$

where $\mathbf{x} = [1, x_1, x_2, \ldots, x_n]^T$

Cost Function

- Measures how well the hypothesis fits the data.
- Mean Squared Error (MSE) cost function:

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

- m = number of training examples.
- Goal: find θ that minimizes $J(\theta)$.

Gradient Descent Algorithm(Click for proof)

- Optimization algorithm to minimize cost function.
- Update rule for each parameter θ_j:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

- For linear regression:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

- α is the learning rate.

Regression Performance Metrics

- **Mean Squared Error (MSE):**

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

- **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\text{MSE}}$$

- **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|h_\theta(x^{(i)}) - y^{(i)}|$$

- **R-squared (R^2):**

$$R^2 = 1 - \frac{\sum(y^{(i)} - h_\theta(x^{(i)}))^2}{\sum(y^{(i)} - \bar{y})^2}$$

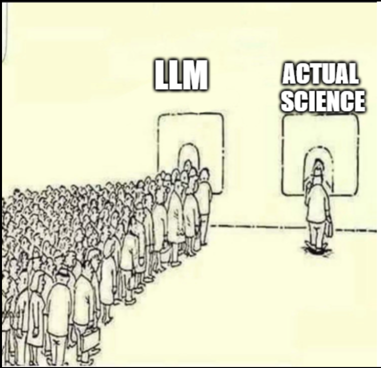
where $\bar{y}$ is the mean of $y^{(i)}$.

Summary

- Supervised learning: regression predicts continuous output.
- Univariate vs multivariate linear regression.
- Hypothesis function models the relationship.
- Cost function measures prediction error.
- Gradient descent optimizes parameters.
- Performance metrics evaluate regression quality.

KEYSIGHT

Week 03- Session 02 - Discrete Event Simulation



Kelvin C. - PassDowns/Walkthroughs III

DES-Intro

DES is a modelling technique where the operation of a system is represented as a chronological sequence of events. Each events occurs at a discrete point in time and changes the systems's state.

Characterisitcs

- Models complex, stochastic systems without oversimplifying
- Tracks entities, events and systems states over time
- Events can include arrival, departure, breakdown, repairs, etc.

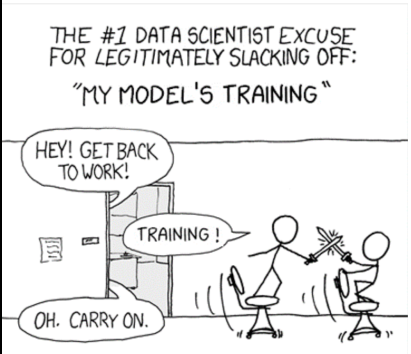
Applications

- Manufacturing production (assembly lines,production line, warehouse inventory)
- Airport operation(baggage handling/routing,check-in systems)
- Logistics and supply chains (warehouse planning)
- Healthcare systems (patient flow in Accident & Emergency Department)

Discrete Event Simulation for Lead Time Forecasting

- **Context:** Manufacturing production line with queues at various stages
- **Goal:** Forecast lead times by modeling the dynamics of work-in-progress
- **What is Discrete Event Simulation (DES)?**
  - A computational method that models the operation of a system as a sequence of discrete events
  - Events correspond to changes in system state, e.g., job arrival, processing start/end, queue departure
- **Why use DES for queuing theory?**
  - Captures stochastic variability in arrivals and processing times
  - Provides detailed insights into queue lengths, queue duration, waiting times, and bottlenecks
  - Enables evaluation of different production scenarios and policies

The end of Walkthroughs......



TBT of Data Scientist's MO.