Lecture 5

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Q-Q Plot (Quantile-Quantile Plot)

Quantile ranges from 0 to 1 ($\sim$ to percentiles).
The plot graphs data quantiles (x) vs theoretical quantiles (z) .

Typically, used for normal distribution, can be extended to other distributions also by changing the z-distribution on the horizontal axis.
If the data is Gaussian-distributed, we will get a straight line on the Q-Q plot.

If the graph cuts the straight line from the x-side, then it is left skewed, else right skewed. This can be seen with the help of the distinguishing compression for $< 50\%ile$ and above them.

Database! What is a Database?

RDBMS (Relational Database Management System)

- Tables
- Records
- Fields
- Keys

SQL (Structured Query Language)

Standard language for interacting with (R)DBMS

- CREATE TABLE
- INSERT INTO
- DELETE FROM

Typical query structure in SQL:

- FROM table
- WHERE condition
- SELECT

ML for Smart Monkeys

ML acts as an estimator by using data to create models that can make new predictions about similar data.

```
Sweet Spot for ML: Lots of Structured Stationary Data
```

Parameters and Hyper-Parameters

- Parameters: variables whose values are updated during training
- Hyperparameters: whose values are fixed by model developer before beginning of learning process.

Types of ML Problems:

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning, Reinforcement Learning, etc.

ML Recipe

- Type of ML problem
- Data Preparation
- ML Framework Selection
- Prepare training, test, validation data
- Perform training, validation and testing

Bias-Variance Trade-Off

- **Underfitting**: High bias, Low variance
- **Overfitting**: Low bias, High variance