1.Differnece between ai and ml.

|  |  |
| --- | --- |
| **AI** | ML |
| Super set of ml | Subset of ai |
| It try to create capable computers that can perform the human task | It try to make algorithm that make predictions and improve its performance based on data |
| Wider scope | Narrower scope |
| It can work with unstructured data | Not work with unstructured data |
| It mimics the human intelligence | Enable machine to learn from data and make predictions. |
| Type:narrow ai,general ai super intelligence AI | Type:supervised,unsupervised and reinforcement learning |
| Example:self driving car | Spam filters |

2. Give advantages and disadvantages of KNN

Answer:

Advantages:

* 1.**Simplicity**: Easy to understand and implement.
* 2. **No Training Phase** KNN is a lazy learner, meaning it doesn't require a training phase. It memorizes the data during training.
* 3. **Versatility**: Suitable for both classification and regression tasks.

Disadvantages:

* **High computational cost:**
* This can lead to poor performance if the dataset contains many irrelevant features.
* **Sensitivity to Noise**: Sensitive to outliers or noisy data, as it relies on distances.
* **Choosing the Right K**: The performance of KNN can be affected by the choice of the number of neighbors (K), and finding the optimal K can be challenging.

3. Describe the process of feature engineering in machine learning.

Answer:

Feature engineering is a crucial step in the machine learning process, involving the transformation of raw data into meaningful features that can be effectively utilized by machine learning algorithms. It plays a pivotal role in enhancing the performance and accuracy of predictive models

Collecting data

Handle missing data

Select best features(input)

4. Explain the concept of overfitting in machine learning.

Answer:

In machine learning, overfitting happens when a model learns not just the general patterns in the data but also the noise or specifics of the training set. As a result, it performs really well on the training data but struggles with new, unseen data because it has essentially memorized the training set rather than understanding the underlying concepts.

5. How does reinforcement learning differ from supervised and unsupervised learning?

Answer:

|  |  |  |
| --- | --- | --- |
| Rainforcement learning | Supervised learning | Un supervised leaning |
| Learns by the interaction with an environment | Learn from label data | Learns from un label data |
| Receives feedback as a punishment or reward | As a input data | No feedback |
| Aims to find best actions among all actions | Aims to map input and output data | Aims to uncover the patterns from raw data |
| Algorithm: Q-learning | Decision tree | Principal component analysis |
| Application :robotic control | Spam detection | clustering |

6. What are training and test data

Answer:

Training data and test data are two essential components of the machine learning process. They play distinct roles in developing and evaluating machine learning models.

**Training Data:**

Training data is a subset of the original dataset that is used to train the machine learning model. It serves as the foundation upon which the model learns to identify patterns, make predictions, or perform specific tasks. The quality and quantity of training data significantly impact the model's performance.

**Test Data:**

Test data, on the other hand, is a separate, unseen portion of the original dataset that is used to evaluate the trained model's performance. It helps assess how well the model generalizes to new data it hasn't encountered during training. This is crucial for ensuring the model's effectiveness in real-world scenarios.

Points:

1. **Purpose:** Training data is used to build the model, while test data is used to assess its performance.
2. **Exposure:** Training data is fed to the model during the learning process, while test data remains unseen until evaluation.

7.Write the differences between Linear and Logistic regression.