MACHINE LEARNING

Assignment 1

by Himanshu Sardana

Contents

Question 1

Given the data points, fit a linear regression model and predict the value for $x = 6$.

```
y = np.array([2, 3, 5, 7, 11])
```

- Use `train_test_split` to split the data into training and testing sets.
- Fit the model using the training set.
- Predict the value for $x = 6$ using the fitted model.

Code

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import numpy as np

# Data points
y = np.array([2, 3, 5, 7, 11])
x = np.arange(1, len(y) + 1).reshape(-1, 1)  # Reshape for sklearn

# Split the data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,
random_state=42)

# Fit the linear regression model
model = LinearRegression()
model.fit(x_train, y_train)

# Predict the value for x = 6
prediction = model.predict(np.array([[6]]))

print(f"Predicted value for x = 6: {prediction[0]}")
```

Output

The predicted value for $x = 6$ is approximately 13.0.

Question 2

Explain the concept of overfitting in machine learning. How can it be prevented?

Solution

Overfitting occurs when a machine learning model learns the training data too well, capturing noise and fluctuations rather than the underlying pattern. This leads to poor generalization to new, unseen data.