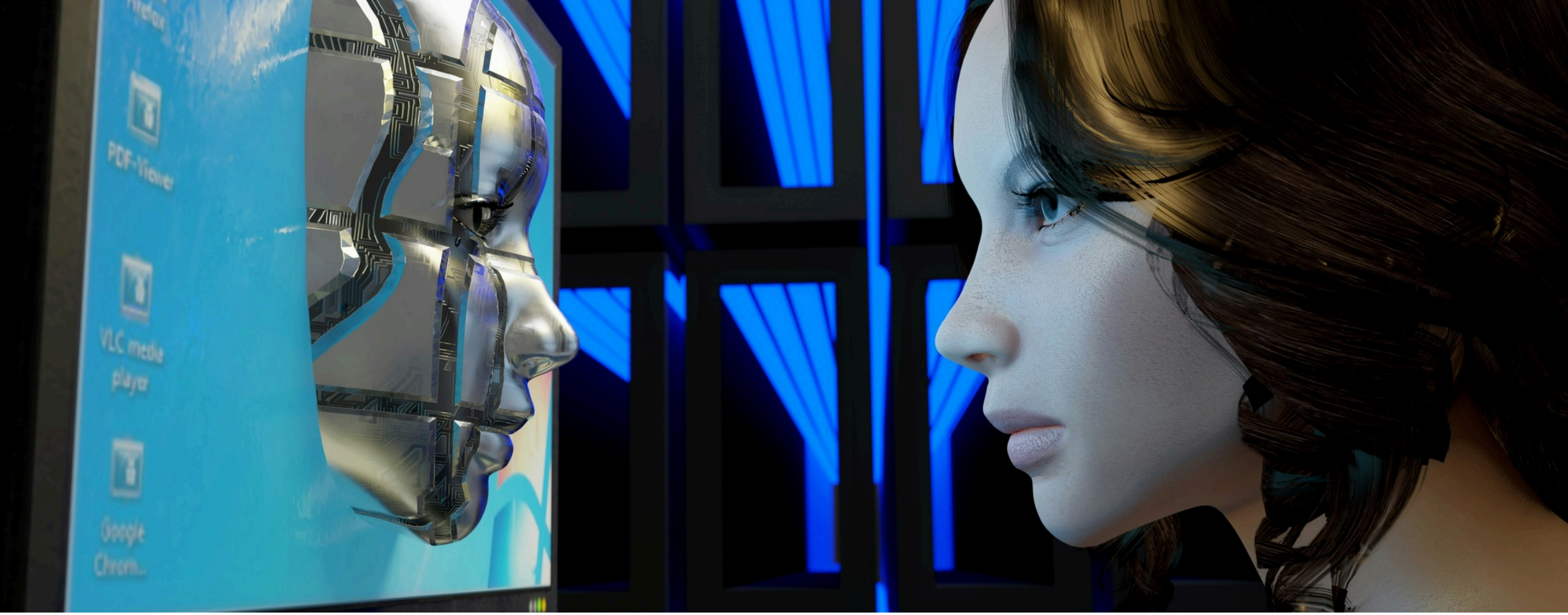


Machine Learning Fundamentals: Linear Regression & Classification Metrics

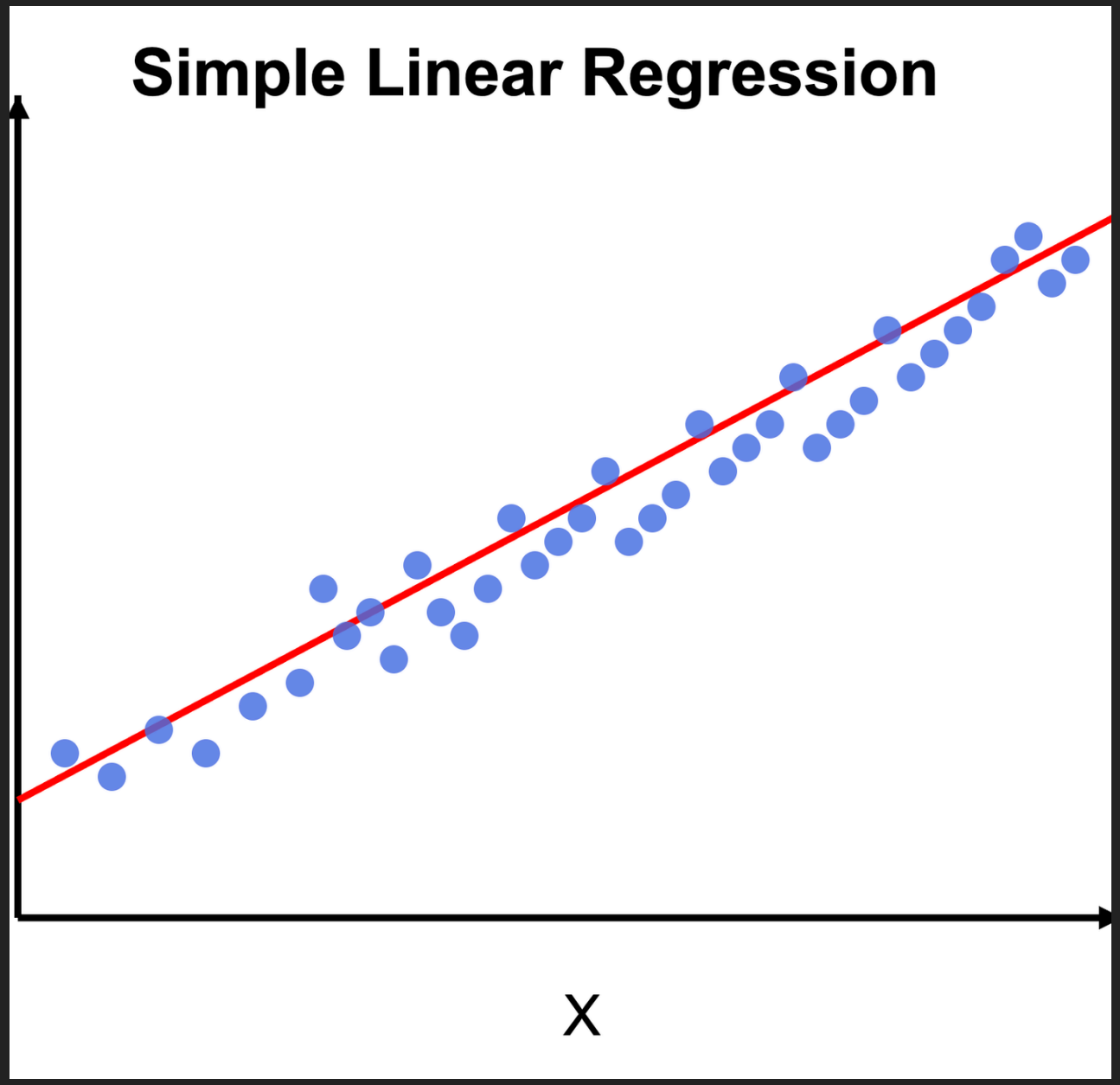
Explore the fundamentals of machine learning, including linear regression and classification metrics.

Linear Regression

- **Import necessary libraries**
  Import Python libraries like NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn for data manipulation, model building, and visualization.
- **Generate a sample dataset**
  Create a synthetic dataset with four input features and a target variable that has a linear relationship with the features, plus some random noise.
- **Split data, scale features, and train the model**
  Split the data into training and testing sets, apply feature scaling, and train a LinearRegression model using the training data.
- **Make predictions and evaluate the model**
  Use the trained model to make predictions on the test set and evaluate its performance using metrics like Mean Squared Error, Mean Absolute Error, and R-squared Score.
- **Visualize results**
  Plot the actual vs. predicted values and the distribution of residuals to gain insights into the model's performance.

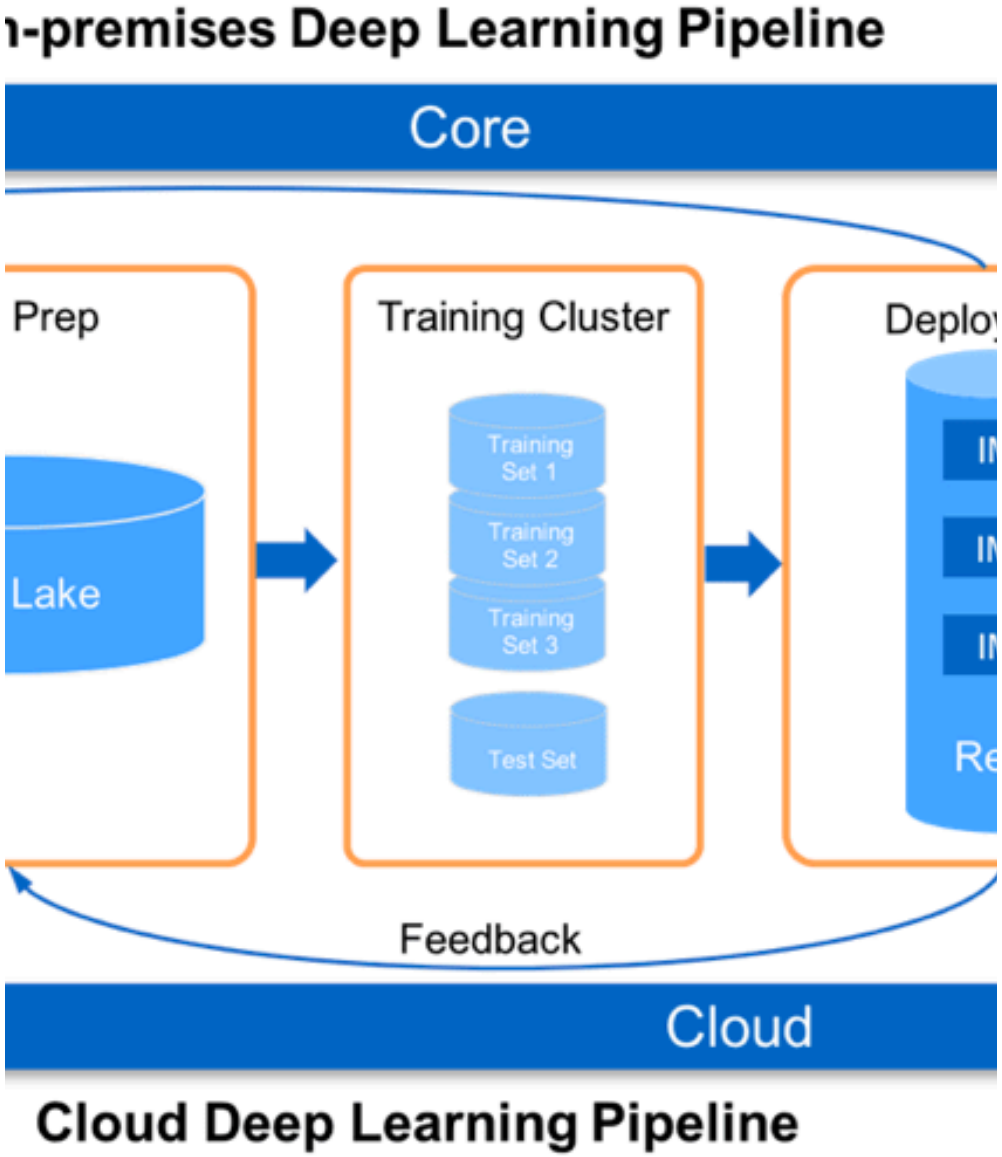
Problem 1: Predicting a Target Value

In this first problem, we'll generate a synthetic dataset to understand the core components of a linear regression model, from data creation to model evaluation. We'll start by importing the necessary libraries, generating a sample dataset with four input features and a target variable that has a linear relationship with the features, plus some random noise. We'll then split the data into training and testing sets, scale the features, and train a linear regression model. Finally, we'll make predictions on the test set and evaluate the model's performance using key metrics like Mean Squared Error, Mean Absolute Error, and R-squared Score.



Problem 2: Salary Prediction

In this problem, we'll predict salary based on categorical and numerical data, introducing preprocessing pipelines to handle mixed data types. We create a dataset containing an employee's 'Domain', 'Experience', 'Education', and 'Salary', with the goal of predicting Salary using the other three features. Since the data contains both categorical and numerical features, we use a preprocessing pipeline to convert the categorical features into a numerical format before feeding them into the linear regression model.



Classification Metrics

Accuracy

The proportion of correct predictions out of all predictions.

Precision

Of all the positive predictions made by the model, how many were actually correct.

Recall (Sensitivity)

Of all the actual positive cases, how many did the model correctly identify.

F1-Score

The harmonic mean of Precision and Recall, providing a single score that balances both.

These key classification metrics help evaluate the performance of a model in predicting categorical outcomes.

Predicting High vs. Low Salary

In this problem, we transform the salary dataset into a classification problem, where the goal is to predict whether an employee has a 'High Salary' or 'Low Salary' based on their domain, experience, and education. We use a logistic regression model to make these predictions and evaluate the model's performance using various classification metrics.

Conclusion

Key Concepts Covered

In this introductory module, you learned about two fundamental concepts in machine learning: Linear Regression and Classification Metrics.

Linear Regression

You built and trained linear regression models to predict continuous target variables, using both synthetic and real-world datasets. You learned how to handle mixed data types, scale features, and evaluate model performance using metrics like MSE, MAE, and R-squared.

Classification Metrics

You transformed the salary dataset into a binary classification problem and learned how to evaluate classification models using metrics like Accuracy, Precision, Recall, and F1-Score. You also explored the use of Confusion Matrices to visualize model performance.

Skills Gained

Through the hands-on exercises in this module, you've gained the following skills:

- Data Preprocessing and Handling

- Building and Training Linear Regression and Logistic Regression Models

- Evaluating Model Performance

- Visualizing and Interpreting Model Results

In this introductory module, you have gained a solid understanding of two fundamental concepts in machine learning: linear regression for predictive modeling and key metrics for evaluating classification models. You have built and evaluated both regression and classification models using practical examples, demonstrating your ability to work with real-world data and apply machine learning techniques. Congratulations on your progress, and we look forward to seeing you explore more advanced topics in the next module!