

```

In [1]: import pandas as pd
from sqlalchemy import create_engine, MetaData, Table, Column, Float, Integer
import numpy as np
from bokeh.plotting import figure, show, output_file
from bokeh.layouts import gridplot
from bokeh.models import Legend
from scipy import stats

# Constants for database connection
DATABASE_URI = 'sqlite:///data.db'

class DataProcessor:
    """Class to handle data processing tasks such as loading data and choosing ideal functions."""

    def __init__(self, db_uri=DATABASE_URI):
        # Initialize database connection
        self.engine = create_engine(db_uri)
        self.metadata = MetaData()

        # Define training data table schema
        self.training_data_table = Table(
            'training_data', self.metadata,
            Column('id', Integer, primary_key=True),
            Column('x', Float),
            Column('y1', Float),
            Column('y2', Float),
            Column('y3', Float),
            Column('y4', Float)
        )

        # Define ideal functions table schema
        self.ideal_functions_table = Table(
            'ideal_functions', self.metadata,
            Column('id', Integer, primary_key=True),
            Column('x', Float),
            *[Column(f'y{i}', Float) for i in range(1, 51)]
        )

        # Create tables if they don't exist
        self.metadata.create_all(self.engine)

    def load_data(self, train_csv_path, test_csv_path, ideal_csv_path):
        """Load data from CSV files into the database."""
        try:
            # Load training data
            train_df = pd.read_csv(train_csv_path)
            train_df.to_sql('training_data', self.engine, if_exists='replace', index=False)
            print("Training data loaded into database.")

            # Load ideal functions data
            ideal_df = pd.read_csv(ideal_csv_path)
            ideal_df.to_sql('ideal_functions', self.engine, if_exists='replace', index=False)
            print("Ideal functions data loaded into database.")

            # Load test data
            test_df = pd.read_csv(test_csv_path)
            print("Test data loaded.")

            return train_df, test_df, ideal_df
        except Exception as e:
            print(f"Error loading data: {e}")
            raise

    def choose_ideal_functions(self, train_df, ideal_df):
        """Choose the ideal functions that best fit the training data using the Least-Square criterion."""
        try:
            chosen_ideal_functions = {}
            for i in range(1, 5):
                y_column = f'y{i}'
                y_train = train_df[y_column].values

                min_error = float('inf')
                ideal_function_index = None

                for j in range(1, 51):
                    ideal_column = f'y{j}'
                    y_ideal = ideal_df[ideal_column].values
                    error = np.sum((y_train - y_ideal) ** 2)

                    if error < min_error:
                        min_error = error
                        ideal_function_index = j

                chosen_ideal_functions[y_column] = (ideal_function_index, min_error)
            print(f"Ideal function for y{i}: y{ideal_function_index} with error {min_error}")

            return chosen_ideal_functions

```

```

    except Exception as e:
        print(f"Error choosing ideal functions: {e}")
        raise

def map_test_data(self, test_df, chosen_ideal_functions, ideal_df):
    """Map test data to chosen ideal functions based on deviation criteria."""
    try:
        # Create a DataFrame to store results
        mapped_df = pd.DataFrame(columns=['x', 'y', 'chosen_ideal_function', 'deviation'])

        for _, row in test_df.iterrows():
            x_test = row['x']
            y_test = row['y']

            # Initialize minimum deviation
            min_deviation = float('inf')
            chosen_ideal_function = None

            # Compare y_test with ideal functions
            for y_column, (ideal_function_index, _) in chosen_ideal_functions.items():
                ideal_column = f'y{ideal_function_index}'
                y_ideal = ideal_df.loc[ideal_df['x'] == x_test, ideal_column].values[0]
                deviation = abs(y_test - y_ideal)

                if deviation < min_deviation:
                    min_deviation = deviation
                    chosen_ideal_function = y_column

            # Append the result
            mapped_df = mapped_df.append({
                'x': x_test,
                'y': y_test,
                'chosen_ideal_function': chosen_ideal_function,
                'deviation': min_deviation
            }, ignore_index=True)

        print("Test data mapped to ideal functions.")
        return mapped_df
    except Exception as e:
        print(f"Error mapping test data: {e}")
        raise

class Visualizer:
    """Class to handle visualization tasks using Bokeh."""

    def plot_data(self, train_df, test_df, ideal_df, mapped_df):
        """Plot training data, chosen ideal functions, test data, and mapped test data."""
        # Create plots
        training_plot = figure(title="Training Data and Ideal Functions", x_axis_label='x', y_axis_label='y')
        test_plot = figure(title="Test Data", x_axis_label='x', y_axis_label='y')
        deviation_plot = figure(title="Mapped Test Data and Deviations", x_axis_label='x', y_axis_label='deviation')

        # Colors for different data sets
        colors = ['blue', 'green', 'red', 'purple']

        # Plot training data
        for i in range(1, 5):
            y_column = f'y{i}'
            training_plot.circle(train_df['x'], train_df[y_column], color=colors[i - 1], size=6, legend_label=f'A')

        # Plot chosen ideal functions
        for i in range(1, 5):
            y_column = f'y{i}'
            ideal_function_index = chosen_ideal_functions[y_column][0]
            ideal_column = f'y{ideal_function_index}'
            training_plot.line(ideal_df['x'], ideal_df[ideal_column], color=colors[i - 1], line_width=2, legend_label=f'B')

        # Plot test data
        test_plot.circle(test_df['x'], test_df['y'], color='black', size=6, legend_label='Test Data')

        # Plot mapped test data and deviations
        for i in range(1, 5):
            y_column = f'y{i}'
            data_subset = mapped_df[mapped_df['chosen_ideal_function'] == y_column]

            # Scatter plot of mapped test data
            test_plot.scatter(data_subset['x'], data_subset['y'], color=colors[i - 1], size=6, legend_label=f'A')

            # Calculate curve fitting line for deviations
            slope, intercept, _, _, _ = stats.linregress(data_subset['x'], data_subset['deviation'])

            # Generate line of best fit for deviations
            x_range = np.linspace(data_subset['x'].min(), data_subset['x'].max(), 100)
            y_fit = slope * x_range + intercept

            # Plot curve fitting line for deviations
            deviation_plot.line(x_range, y_fit, color=colors[i - 1], line_width=2, legend_label=f'Deviation fit')
            deviation_plot.scatter(data_subset['x'], data_subset['deviation'], color=colors[i - 1], size=6, legend_label=f'A')

```

```
# Configure legends
training_plot.legend.title = 'Data Types'
training_plot.legend.title_text_font_style = 'bold'
training_plot.legend.location = 'top_left'

test_plot.legend.title = 'Test Data and Mapped Test Data'
test_plot.legend.title_text_font_style = 'bold'
test_plot.legend.location = 'top_left'

deviation_plot.legend.title = 'Mapped Test Data and Deviations'
deviation_plot.legend.title_text_font_style = 'bold'
deviation_plot.legend.location = 'top_left'

# Combine plots in a grid
layout = gridplot([training_plot], [test_plot], [deviation_plot])

# Save and show the plots
output_file("data_visualization.html")
show(layout)

if __name__ == "__main__":
    # File paths
    train_csv_path = "C:\\Users\\sanja\\OneDrive\\Desktop\\IU\\train.csv"
    test_csv_path = "C:\\Users\\sanja\\OneDrive\\Desktop\\IU\\test.csv"
    ideal_csv_path = "C:\\Users\\sanja\\OneDrive\\Desktop\\IU\\ideal.csv"

    # Create an instance of DataProcessor
    data_processor = DataProcessor()

    # Load data from CSV files
    train_df, test_df, ideal_df = data_processor.load_data(train_csv_path, test_csv_path, ideal_csv_path)

    # Choose ideal functions
    chosen_ideal_functions = data_processor.choose_ideal_functions(train_df, ideal_df)

    # Map test data to chosen ideal functions
    mapped_df = data_processor.map_test_data(test_df, chosen_ideal_functions, ideal_df)

    # Create an instance of Visualizer
    visualizer = Visualizer()

    # Plot data
    visualizer.plot_data(train_df, test_df, ideal_df, mapped_df)
```

```

Training data loaded into database.
Ideal functions data loaded into database.
Test data loaded.
Ideal function for y1: y42 with error 34.246594303368504
Ideal function for y2: y41 with error 35.60184692481152
Ideal function for y3: y11 with error 29.86183029016382
Ideal function for v4: v48 with error 31.963434327891697

```

[illegible]

[illegible]

[illegible]


```
mapped_at = mapped_at.append(
    Test data mapped to ideal functions.
```