

✓ 3(a)

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('poly_data.csv', names=['X', 'Y'], header=None)
X = data[['X']].values
y = data['Y'].values

from sklearn.model_selection import KFold
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

degrees = range(1, 41)
k = 5
kf = KFold(n_splits=k, shuffle=True, random_state=69)

cv_errors = []

for degree in degrees:
    poly = PolynomialFeatures(degree)
    X_poly = poly.fit_transform(X)

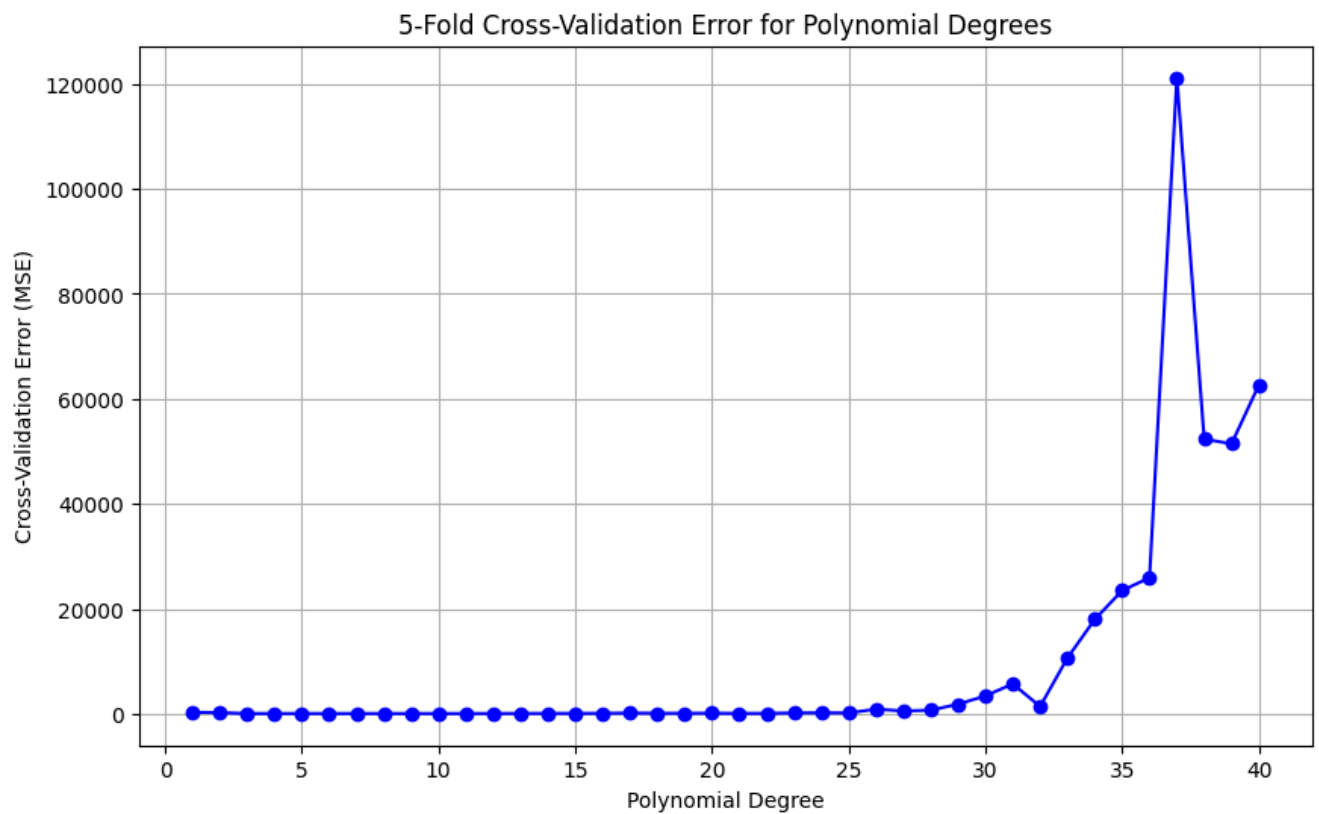
    fold_errors = []

    for train_index, val_index in kf.split(X_poly):
        X_train, X_val = X_poly[train_index], X_poly[val_index]
        y_train, y_val = y[train_index], y[val_index]
        # print(len(X_train), len(y_train))
        model = LinearRegression()
        model.fit(X_train, y_train)

        y_val_pred = model.predict(X_val)
        fold_error = mean_squared_error(y_val, y_val_pred)
        fold_errors.append(fold_error)

    avg_cv_error = np.mean(fold_errors)
    cv_errors.append(avg_cv_error)

plt.figure(figsize=(10, 6))
plt.plot(degrees, cv_errors, marker='o', linestyle='-', color='b')
plt.xlabel('Polynomial Degree')
plt.ylabel('Cross-Validation Error (MSE)')
plt.title(f'{k}-Fold Cross-Validation Error for Polynomial Degrees')
plt.grid(True)
plt.show()
```



```
best_degree = degrees[np.argmin(cv_errors)]  
print(f"The best polynomial degree based on {k} cross-validation is: {best_degree}")
```



The best polynomial degree based on 5 cross-validation is: 9

✓ 3(b)

```
best_degree = degrees[np.argmin(cv_errors)]  
  
poly = PolynomialFeatures(best_degree)  
X_poly = poly.fit_transform(X)  
model = LinearRegression()  
model.fit(X_poly, y)  
  
print(f"Polynomial coefficients for degree {best_degree}:")  
print(model.coef_)  
  
X_curve = np.linspace(X.min(), X.max(), 100).reshape(-1, 1)  
X_curve_poly = poly.transform(X_curve)  
y_curve = model.predict(X_curve_poly)  
  
plt.figure(figsize=(10, 6))  
plt.scatter(X, y, color='blue', label='Data points')  
plt.plot(X_curve, y_curve, color='red', linewidth=2, label=f'Fitted polynomial (degree {best_degree})')  
plt.xlabel('X')  
plt.ylabel('Y')  
plt.title(f'Polynomial Regression Fit (Degree {best_degree})')  
plt.legend()  
plt.grid(True)  
plt.show()
```



Polynomial coefficients for degree 9:

```
[ 0.          -6.32550987  7.72580525 15.2293662  -3.78550145 -5.20498237  
 0.85899842  0.80888092 -0.05668653 -0.04139501]
```

