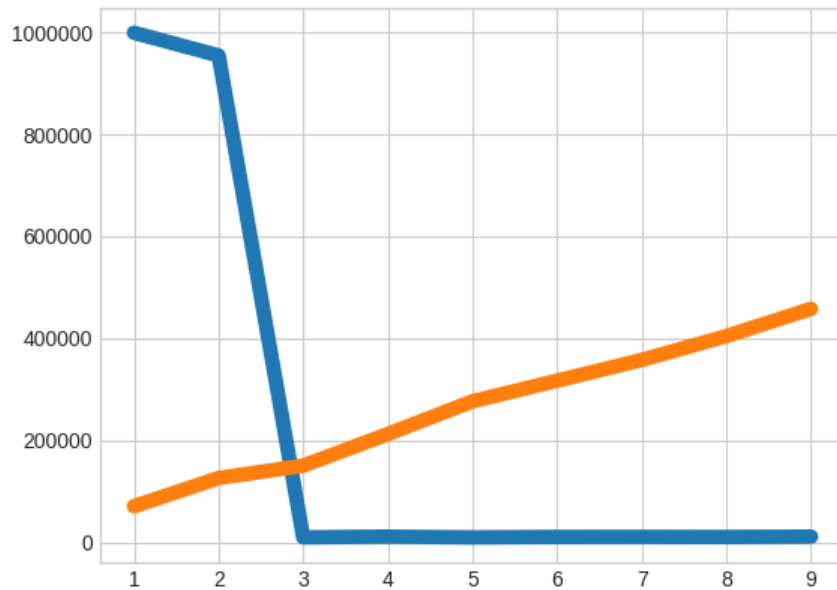


MDL Assignment Report

Q2

Abinash Maharana
Shivansh Anand



X Axis – Model complexity

Y Axis – Error

Blue Line – bias²

Orange - Variance

Observations:

Underfitting:

Bias² is significantly high for degrees 1-2 and variance is low.

Thus we can safely say that the model is underfit when degree is 2 or lower.

Overfitting:

Variance increases steadily as degree increases and bias² becomes very low above degree 3. Maybe we can say the model is overfitting after degree 5-6 but still there is no significant spike in variance.

WE CAN SAFELY SAY THAT THE BEST FIT FOR THIS DATA IS DEGREE 3 OR 4. (MINIMA OF MSE WILL LIE SOMEWHERE AMONG THESE TWO)

Logic:

Training data was divided into 20 parts.

Model was developed on various degrees.

Bias² and variance were plotted.

Code for model fitting:

```
for x in range(1,10):
    poly=PolynomialFeatures(degree=x)
    X_test=poly.fit_transform(Xtest)
    m=linear_model.LinearRegression()
    ll=np.zeros(shape=(len(Xtest),1))
    tpl=[]
    tpl.append(x)
    for y in range(len(xt)):
        x_ttrain=xt[y]
        y_train=yt[y]
        x_ttrain=x_ttrain.reshape(-1,1)
        X_train=poly.fit_transform(x_ttrain)
        m.fit(X_train,y_train)
        tpp=m.predict(X_test)
        tpp=tpp.reshape(-1,1)
        ll=np.hstack((ll,tpp))
```

fit() is used for fitting data to a model using linear regression.

fit_transform() is used to setup training data according to the degree of our model

predict() is used then for bias and variance calculations

reshape() is to get the data into proper 1xn matrix format.