Important codings

* To fill the missing values with mean

for col, value in df.items():

if col != ‘type’: (this oly applies whne there is a string value)

df[col] = df[col].fillna(df[col].mean())

# create box plots

fig, ax = plt.subplots(ncols=6, nrows=2, figsize=(20,10))

index = 0

ax = ax.flatten()

for col, value in df.items():

if col != 'types':

sns.boxplot(y=col,data=df,ax=ax[index])

index+= 1

plt.tight\_layout(pad=0.5,w\_pad=0.7,h\_pad=5.0)

Missing Values

From sklearn.preprocessing import SimpleImputer

* Imp=SimpleImputer(missing\_values=0,strategy=’mean,axis=0)
* Imp=imp.fit(df)
* dst=imp.transform(df.values)
* df = pd.DataFrame(dst)
* df.head()

**Visualization for linear Regression**

Sns.regplot(x=’highway’(independent variable),y=”price(target variable)”,data=df)

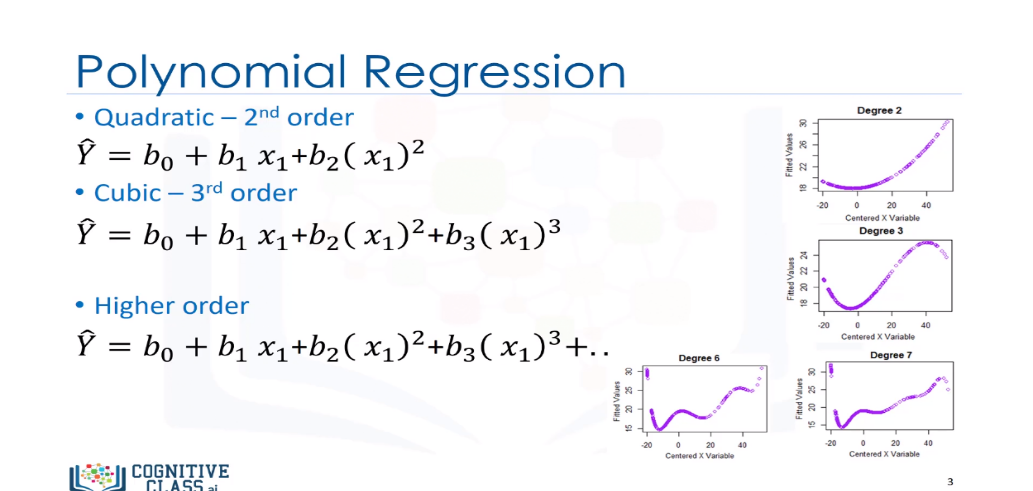
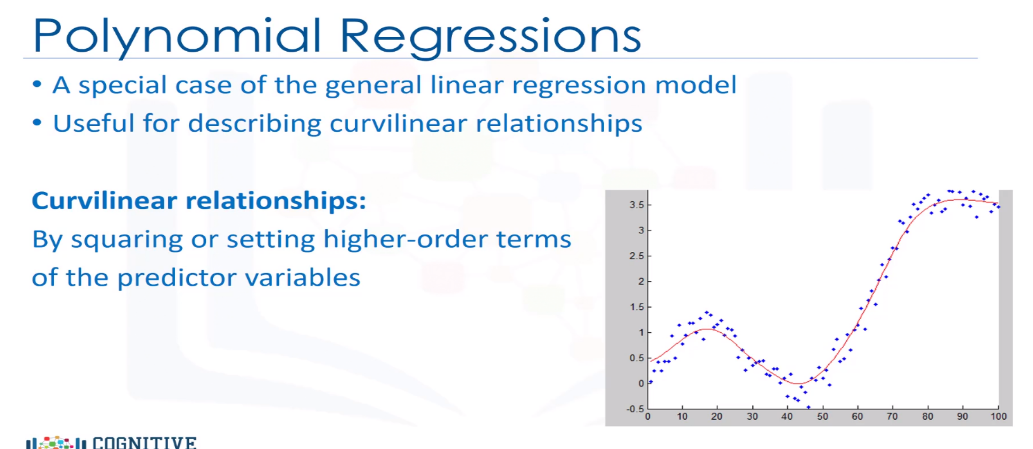
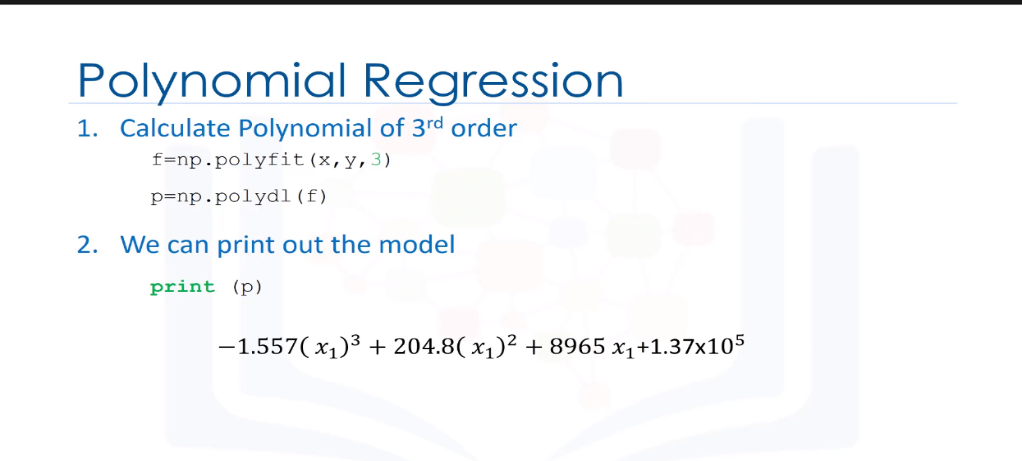
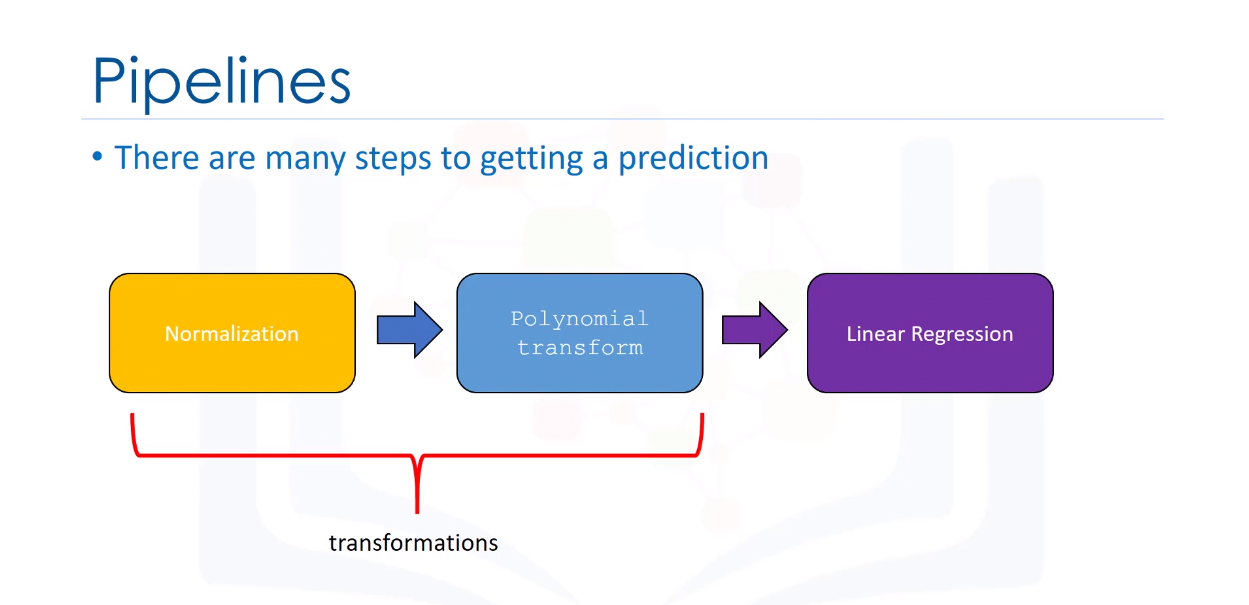
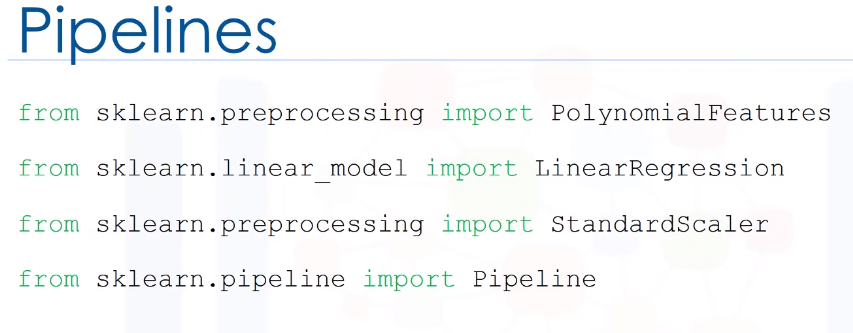
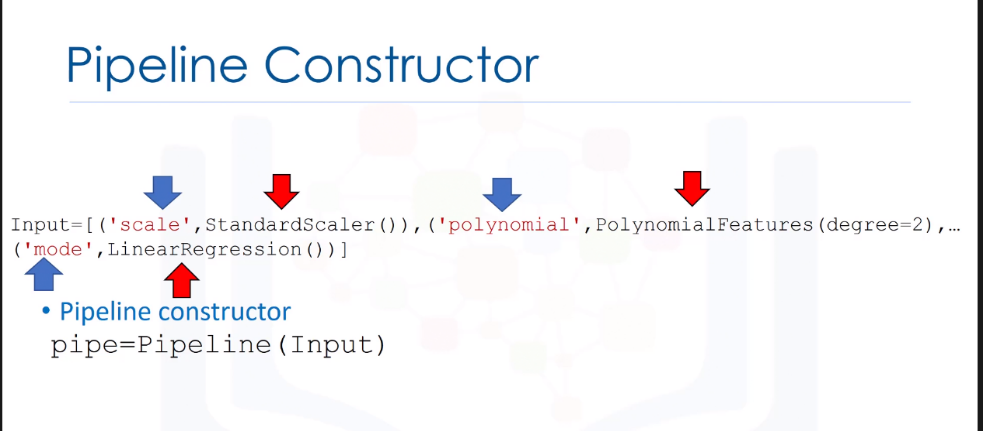
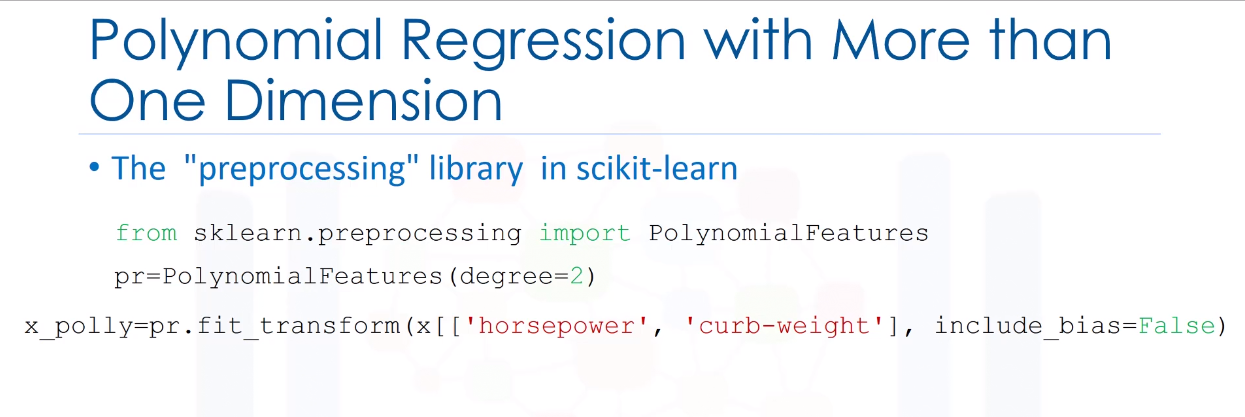
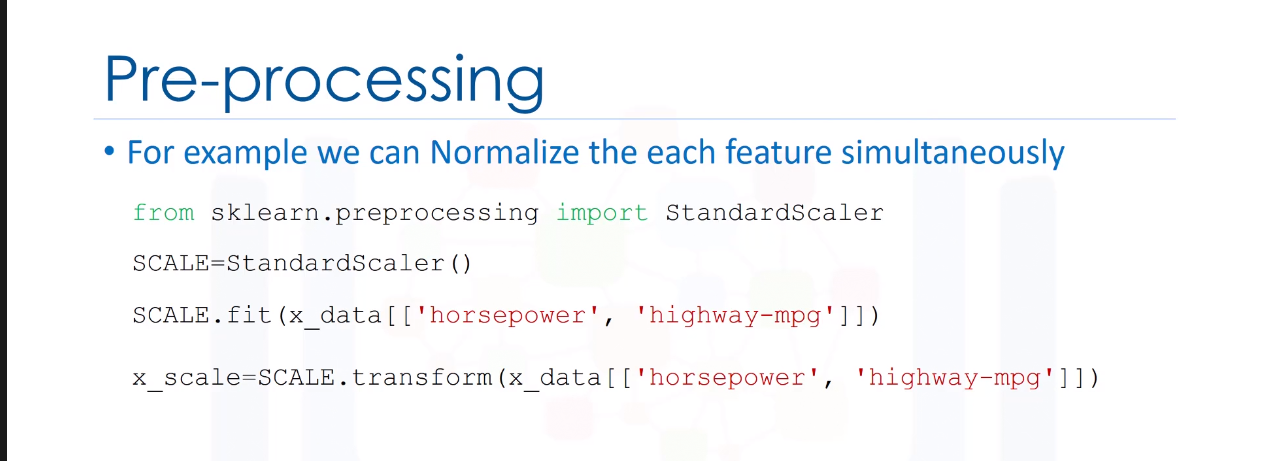
Plt.ylim(0,)

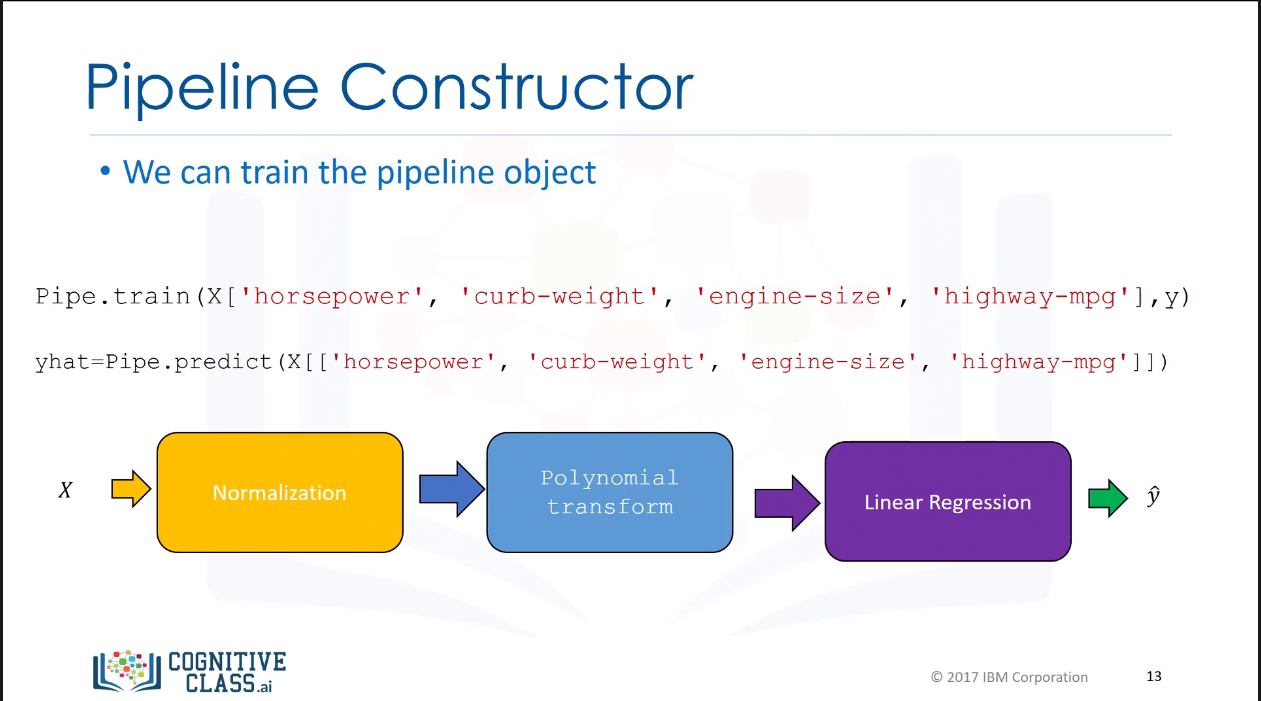
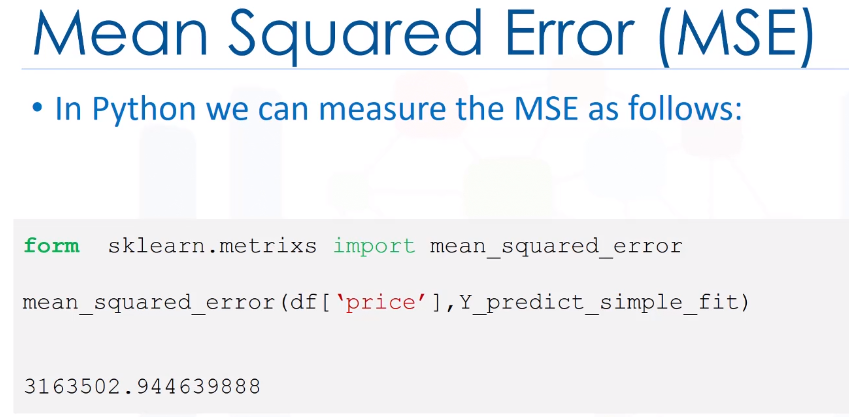
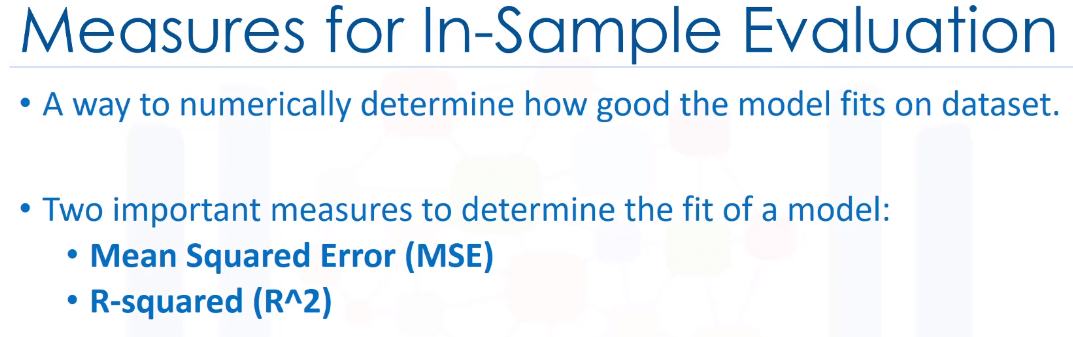
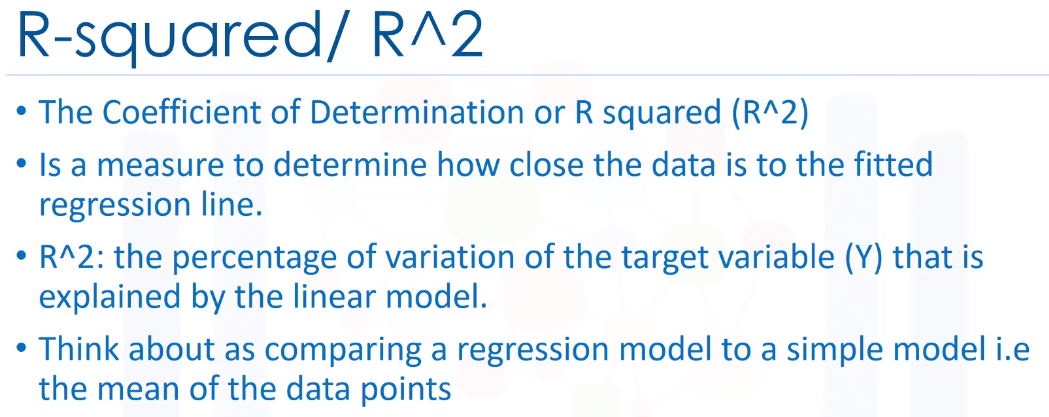
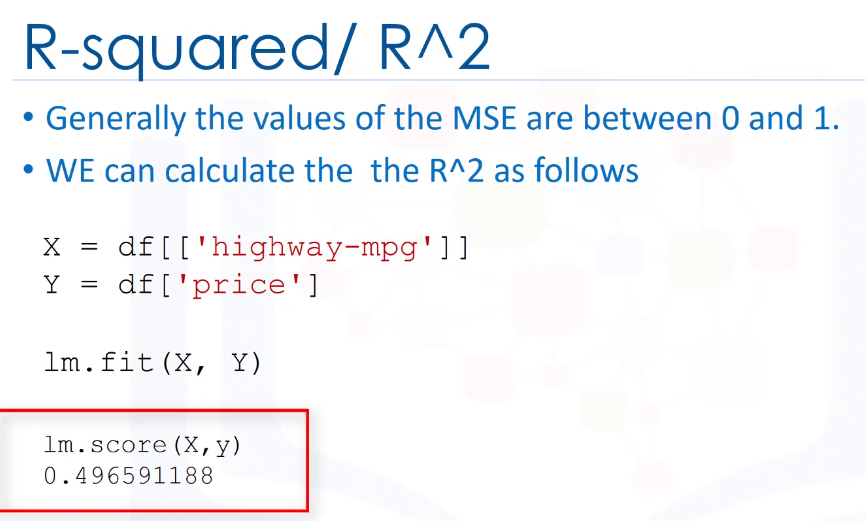
Residual Plot

* difference between the actual value
* Sns.residplot(df[‘hghway],df[‘price’])



Polynominal Regresssion and Pipelines

When the linear model is not the best fit for our model (we transfer out data to polynominal and then we will use linear regression)  

* R-square is always between 0 nd 1 and if it’s a negative value it could be due to overfitting.

