From the "pair plot" visualization, we observe that :

- petal length and petal width are most positively correlated as we see a linear increase between the feaures. The scatter plot aligns with a linear line function.
- we observe a similar pattern with petal length and sepal length where there is linear positive correlation.
- In all the plots, Iris-setosa is easily distinguishable and can be identified irrespective of petal or sepal features. By using petal length, we can distinctly seperate Iris-setosa.
- For Iris-versicolor and Iris-virginica, we see that the plots are mostly overlapping, but petal features provide better distinction than sepal features.

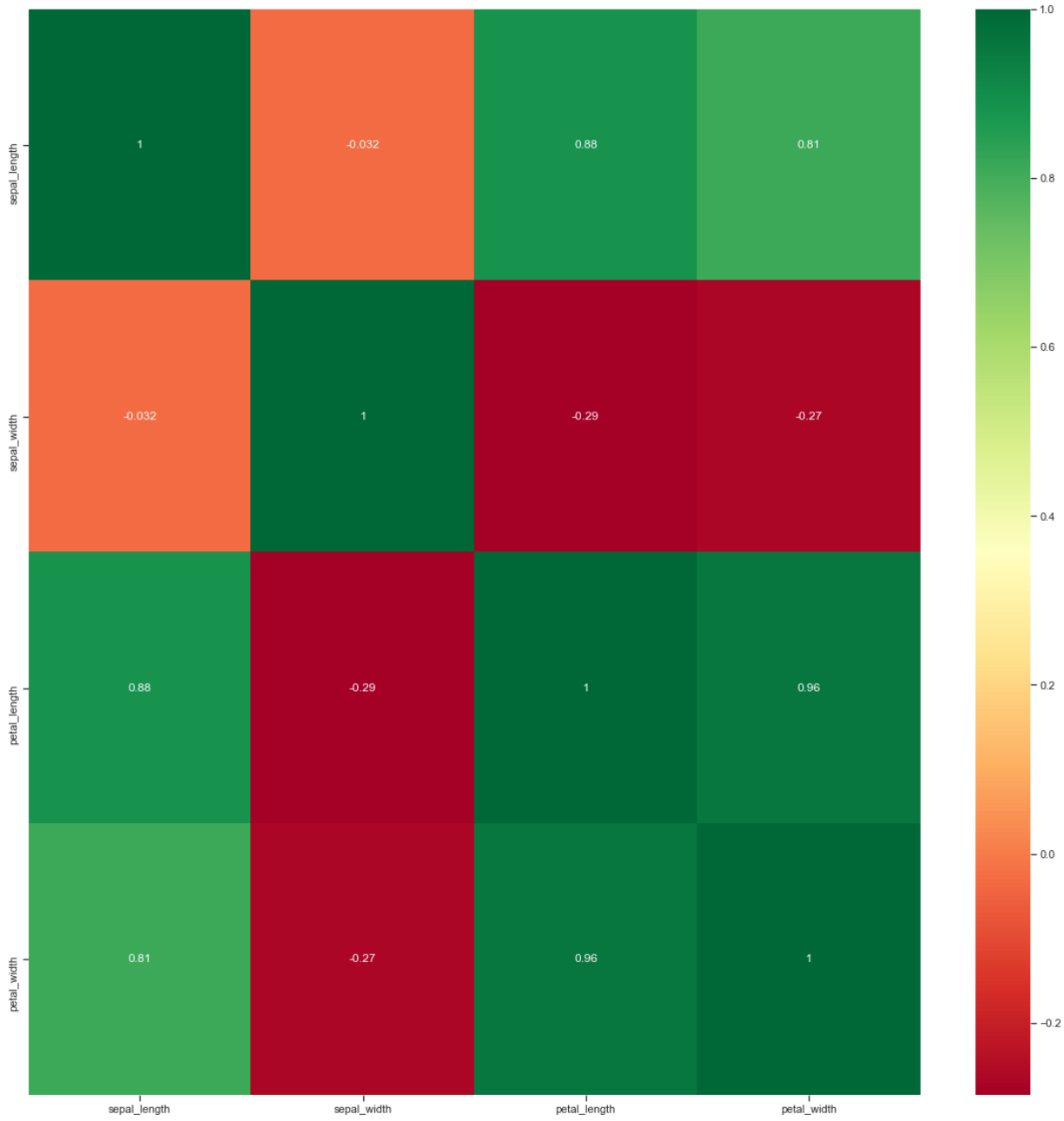
[CM2]

Correlation coefficient of each pair of features

Heat map is used to find out the correlation between different features in the dataset. High positive or negative value shows that the features have high correlation

[7]:
```
# Get correlations of each features in dataset
corrmat = df_iris.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))

# Plot heat map
iris_heat_map=sns.heatmap(df_iris[top_corr_features].corr(),annot=True,cmap='RdYlGn')
```



From the "heat map" visualization, we observe that:

- there is high positive correlation between petal width and petal length i.e increase in petal width corresponds to increase in petal length, and vice versa.
- there is high positive correlation between petal length and sepal length i.e increase in petal length corresponds to increase in petal length, and vice versa.
- there is high positive correlation between petal width and sepal length i.e increase in petal width corresponds to increase in sepal length, and vice versa.
- we see minor negative correlation between petal length, petal width and sepal width.

Mostly we see correlations between the features petal width, petal length and sepal length.As observed in "Pair Plots" we can see petal features are more correlated than sepal features.

Calculation of mean, variance, skew, kurtosis for the datasets

[8]:
```
# calculate skew
df_iris.skew()
```

[8]:
```
sepal_length    0.401506
sepal_width     0.367708
petal_length   -0.255767
petal_width    -0.074751
dtype: float64
```

[9]:
```
# calculate kurtosis
df_iris.kurtosis()
```

[9]:
```
sepal_length   -0.544820
sepal_width     0.510490
petal_length   -1.389810
petal_width    -1.315451
dtype: float64
```

[10]:
```
# calculate mean
df_iris.mean()
```

[10]:
```
sepal_length    5.858909
sepal_width     3.059083
petal_length    3.812370
petal_width     1.199708
dtype: float64
```

[11]:
```
# calculate variance
df_iris.var()
```

[11]:
```
sepal_length    0.742420
sepal_width     0.207131
petal_length    3.216602
petal_width     0.619672
dtype: float64
```

From mean, variance, skew, kurtosis of the dataset, we observe that:

- the data is fairly symmetric as the skewness ranges within [-0.5,0.5]
- the kurtosis values for sepal length, petal length and petal width are negative indicating light tail distribution. Whereas, sepal width has positive kurtosis value indicating heavy tail distribution.
- variance for petal length is high indicating that the values are highly spread out from the mean.
- mean is higher for sepal features than petal features,indicating flower sepals are broader and lengthier than the petals.