LAB 5 - Classification II (Logistic Regression)

SUBMITTED BY:

Name : Kavitha.S
Reg no : 21122033
Class : 2MSDS

LAB OVERVIEW:

Apply Logistic Regression for Breast Cancer Dataset.
Use 60:40 train-test ratio for splitting the dataset.

Questions:

1. Demonstrate the Logistic Regression for different penalties/regularisation methods - none, l1, l2 (you may use 'saga' solver as the parameter)
2. What happens when the Maximum Iterations are kept as 1, 2, 5, 10, 20, 50, 100, 500 and 1000? Is there any change in the accuracy.
3. Get the attributes: classes_, coef_ and intercept_ and print the same in the above case.

PROBLEM DEFINITION:

Perform Logistic regression on Breast cancer dataset based on various parameters.

APPROACH:

Imported the dataset and the necessary libraries. Converted the target variable into binary values. Performed EDA on the dataset and normalized the dataset using Standard Scaler. Split the dataframe into train and test set. Performed Logistic Regression with various penalty and iterations and compared the accuracy score of each model.

In [1]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import linear_model
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
from sklearn.linear_model import ElasticNet
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
cancer = pd.read_csv('data.csv')
```

In [3]:
```
cancer.head()
```

Out[3]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mea
0	842302	M	17.99	10.38	122.80	1001.0	0.1184
1	842517	M	20.57	17.77	132.90	1326.0	0.0847
2	84300903	M	19.69	21.25	130.00	1203.0	0.1096
3	84348301	M	11.42	20.38	77.58	386.1	0.1425
4	84358402	M	20.29	14.34	135.10	1297.0	0.1003

5 rows × 33 columns

In [4]:
```
list(cancer.columns)
```

```
'perimeter_se',
'area_se',
'smoothness_se',
'compactness_se',
'concavity_se',
'concave points_se',
'symmetry_se',
'fractal_dimension_se',
'radius_worst',
'texture_worst',
'perimeter_worst',
'area_worst',
'smoothness_worst',
'compactness_worst',
'concavity_worst',
'concave points_worst',
'symmetry_worst',
'fractal_dimension_worst',
'Unnamed: 32']
```

In [5]:
```
cancer = cancer.drop(['Unnamed: 32','id'], 1)
```

In [6]:
```
1 cancer.head()
```

Out[6]:

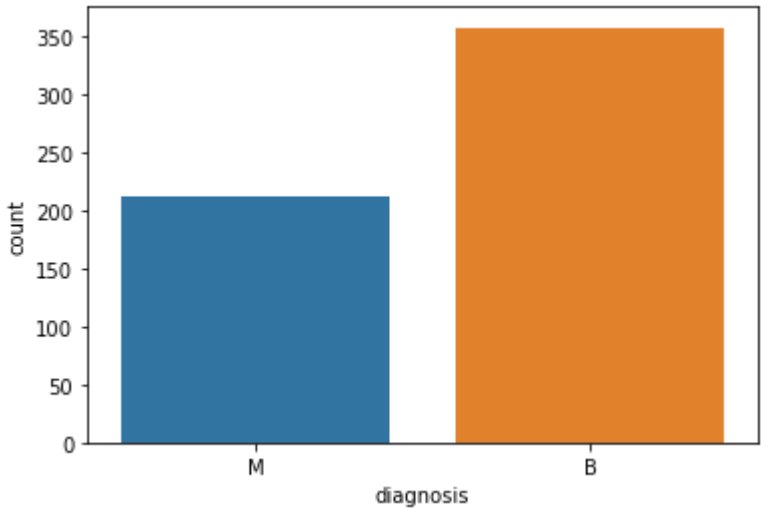
	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compact
0	M	17.99	10.38	122.80	1001.0	0.11840	
1	M	20.57	17.77	132.90	1326.0	0.08474	
2	M	19.69	21.25	130.00	1203.0	0.10960	
3	M	11.42	20.38	77.58	386.1	0.14250	
4	M	20.29	14.34	135.10	1297.0	0.10030	

5 rows × 31 columns

In [7]:
```
1 sns.countplot(cancer['diagnosis'],label='count')
```

```
C:\Users\SRIDHAR\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureW
arning: Pass the following variable as a keyword arg: x. From version 0.12, the
only valid positional argument will be `data`, and passing other arguments with
out an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[7]: <AxesSubplot:xlabel='diagnosis', ylabel='count'>

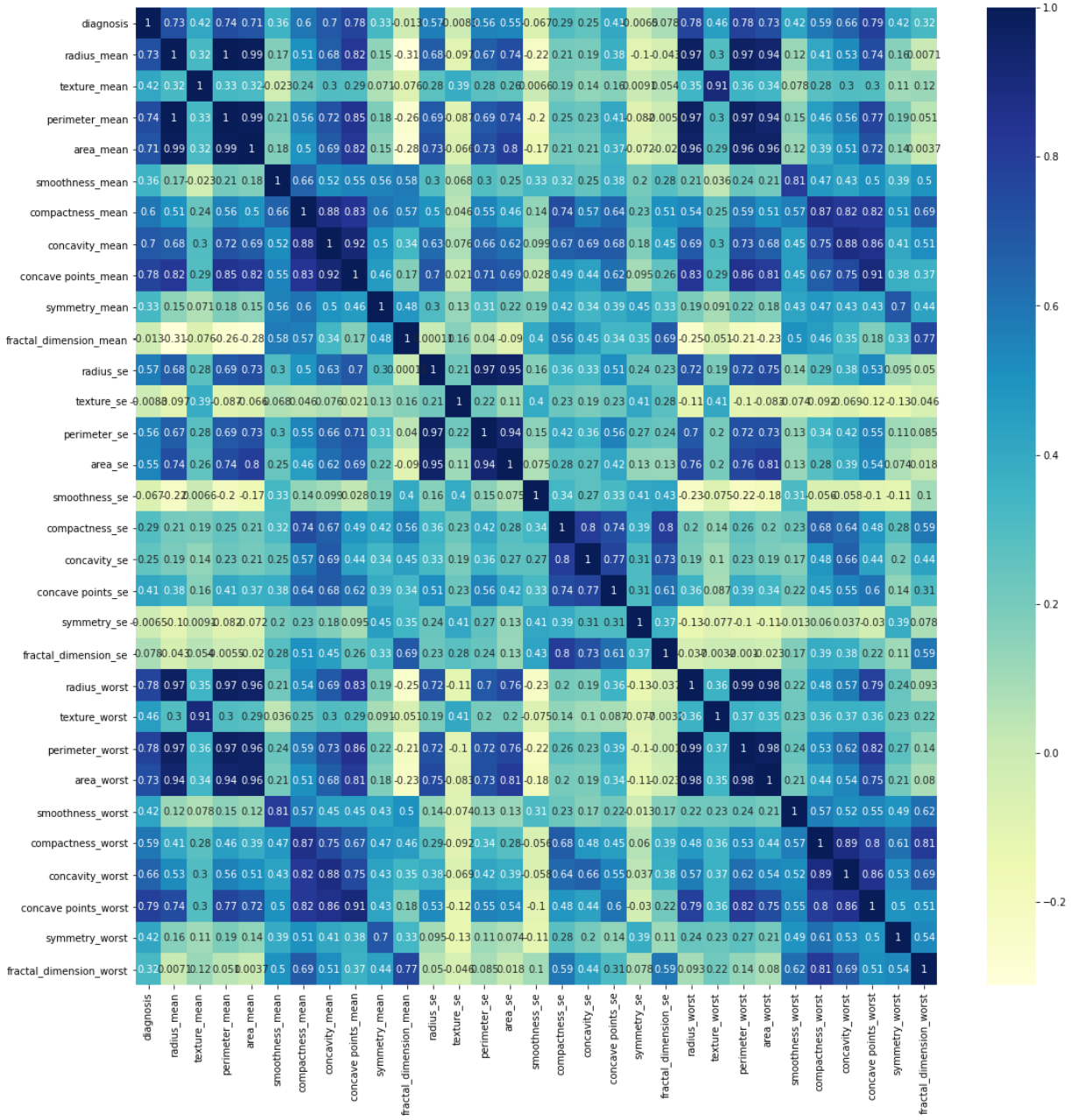


In [8]:
```
1 cancer.diagnosis = [1 if each == "M" else 0 for each in cancer.diagnosis]
```

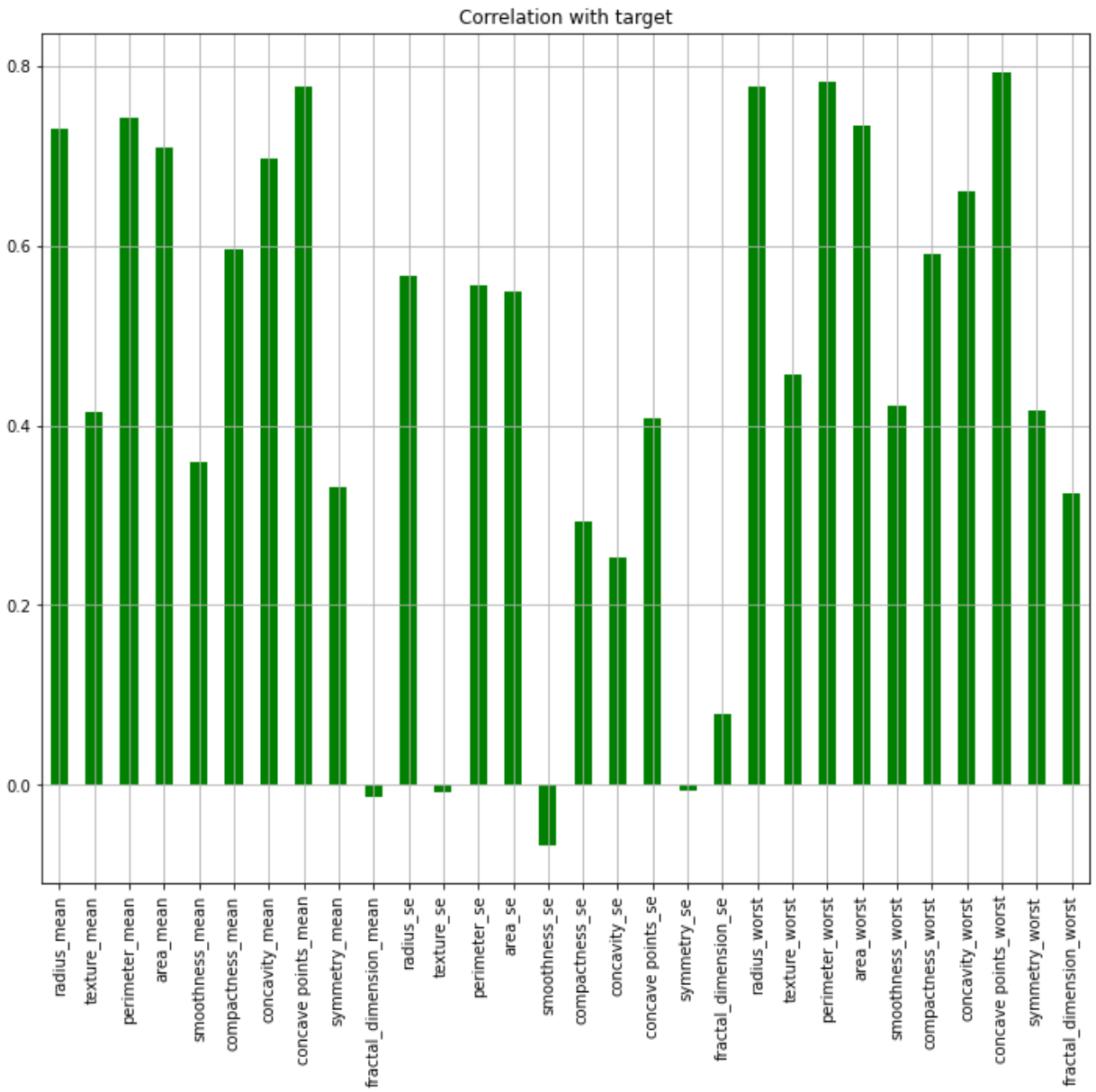
In [9]:
```
1 X = cancer.drop(['diagnosis'],axis=1).values
2 y = cancer['diagnosis'].values
```

In [10]:
```
f,ax = plt.subplots(figsize=(18, 18))
sns.heatmap(cancer.corr(), cmap="YlGnBu",ax=ax,annot= True)
```

Out[10]: <AxesSubplot:>



In [11]:
```
1 cancer.drop('diagnosis', axis=1).corrwith(cancer.diagnosis).plot(kind='bar',
```



In [12]:
```
scaler = StandardScaler()
print(scaler.fit_transform(X))
```

```
[[ 1.09706398 -2.07333501  1.26993369 ...  2.29607613  2.75062224
   1.93701461]
 [ 1.82982061 -0.35363241  1.68595471 ...  1.0870843  -0.24388967
   0.28118999]
 [ 1.57988811  0.45618695  1.56650313 ...  1.95500035  1.152255
   0.20139121]
 ...
 [ 0.70228425  2.0455738   0.67267578 ...  0.41406869 -1.10454895
  -0.31840916]
 [ 1.83834103  2.33645719  1.98252415 ...  2.28998549  1.91908301
   2.21963528]
 [-1.80840125  1.22179204 -1.81438851 ... -1.74506282 -0.04813821
  -0.75120669]]
```

In [13]:
```
X_train, X_test, y_train, y_test = train_test_split( X, y, test_size = 0.4,
```

In [14]:
```
print("X train: ", X_train.shape)
print("X test: ", X_test.shape)
print("y train: ", y_train.shape)
print("y test: ", y_test.shape)
```

```
X train:  (341, 30)
X test:  (228, 30)
y train:  (341,)
y test:  (228,)
```

In [15]:
```
log_reg = LogisticRegression(penalty='l1',solver='saga').fit(X_train, y_trai
y_pred = log_reg.predict(X_test)
print ("Accuracy : ", accuracy_score(y_test, y_pred))
print('class',log_reg.classes_)
print('coef',log_reg.coef_)
print('intercept',log_reg.intercept_)
```

```
Accuracy :  0.9517543859649122
class [0 1]
coef [[-2.30933215e-03 -4.18200743e-03 -1.39223475e-02 -9.85012146e-03
  -1.66180348e-05  0.00000000e+00  1.45030745e-05  2.63743787e-06
  -3.75929438e-05 -1.07676392e-05 -4.82096350e-06 -3.12373845e-04
  -3.43480968e-05  5.43821386e-03  0.00000000e+00  0.00000000e+00
   0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00
  -2.32308578e-03 -5.26666794e-03 -1.38465013e-02  1.09396313e-02
  -2.38116790e-05  2.06707989e-06  3.16003861e-05  9.45101118e-07
  -5.40777719e-05 -1.22571948e-05]]
intercept [-0.00030355]

C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:328:
ConvergenceWarning: The max_iter was reached which means the coef_ did not conv
erge
  warnings.warn("The max_iter was reached which means "
```

In [16]:
```
penalty_list = ['none', 'l1', 'l2']
for i in range(3):
    log_reg = LogisticRegression(penalty=penalty_list[i],solver='saga').fit(
    y_pred = log_reg.predict(X_test)
    print ("Accuracy score for penalty ",penalty_list[i],':', accuracy_score
```

```
Accuracy score for penalty  none : 0.9517543859649122
Accuracy score for penalty  l1 : 0.9517543859649122
Accuracy score for penalty  l2 : 0.9517543859649122

C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:328:
ConvergenceWarning: The max_iter was reached which means the coef_ did not conv
erge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:328:
ConvergenceWarning: The max_iter was reached which means the coef_ did not conv
erge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:328:
ConvergenceWarning: The max_iter was reached which means the coef_ did not conv
erge
  warnings.warn("The max_iter was reached which means "
```

In [18]:

```
result = pd.DataFrame(columns = ['Penalty','Iteration','Accuracy','Class','C
penalty_list = ['none', 'l1', 'l2']
iterations = [1, 2, 5, 10, 20, 50, 100, 500 , 1000]
for i in range(3):
    for j in range(9):
        log_reg = LogisticRegression(penalty=penalty_list[i],solver='saga',m
        y_pred = log_reg.predict(X_test)
        result.loc[len(result)]=[penalty_list[i],iterations[j],accuracy_scor
result
```

```
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
```

```
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
```

```
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
```

Out[18]:

	Penalty	Iteration	Accuracy	Class	Coef	Intercept
0	none	1	0.350877	[0, 1]	[[-5.560256694460249e-05, -9.76281028511887e-0...	[-7.08296492477462e-06]
1	none	2	0.355263	[0, 1]	[[-0.00011179729844872361, -0.0002097597211203...	[-1.3992451660886552e-05]
2	none	5	0.793860	[0, 1]	[[-0.00022087629398740541, -0.0003972172870412...	[-2.737295801595736e-05]
3	none	10	0.925439	[0, 1]	[[-0.00039288322390086515, -0.0007129540932707...	[-4.917161508131445e-05]
4	none	20	0.951754	[0, 1]	[[-0.0006846636059507385, -0.00125012862329360...	[-8.690368468178016e-05]
5	none	50	0.951754	[0, 1]	[[-0.001414823907061922, -0.002582245907792407...	[-0.00018236543036896367]
6	none	100	0.951754	[0, 1]	[[-0.0023199278560362825, -0.00418606124965923...	[-0.00030408938584685334]
7	none	500	0.956140	[0, 1]	[[-0.0057911250934972874, -0.00884292908106907...	[-0.000790382142991931]
8	none	1000	0.964912	[0, 1]	[[-0.007918463733671317, -0.009639219308608659...	[-0.0010993450394378831]
9	l1	1	0.350877	[0, 1]	[[-6.153065836161691e-05, -0.00011003441452732...	[-7.5216307057053565e-06]
10	l1	2	0.451754	[0, 1]	[[-0.00011079232851754575, -0.0001938994650076...	[-1.3726064245645435e-05]
11	l1	5	0.723684	[0, 1]	[[-0.00022481776482399226, -0.0004140486768488...	[-2.791645925751729e-05]
12	l1	10	0.925439	[0, 1]	[[-0.0003917179349103834, -0.00071332179850995...	[-4.909711841580393e-05]
13	l1	20	0.951754	[0, 1]	[[-0.0006891568721552233, -0.00126142669370517...	[-8.731588401944254e-05]
14	l1	50	0.951754	[0, 1]	[[-0.001410721211619074, -0.002575647685541485...	[-0.0001822009325905026]
15	l1	100	0.951754	[0, 1]	[[-0.002311584369831389, -0.004170613544832519...	[-0.00030377576131108335]
16	l1	500	0.956140	[0, 1]	[[-0.005754730993298763, -0.008808033110759806...	[-0.0007909264447219218]
17	l1	1000	0.964912	[0, 1]	[[-0.007845322578171582, -0.009587586329128655...	[-0.001100518208580118]
18	l2	1	0.350877	[0, 1]	[[-6.218829676332364e-05, -0.00011294672214779...	[-7.795647539943972e-06]
19	l2	2	0.381579	[0, 1]	[[-0.0001116061487525492, -0.00020124077198843...	[-1.37487333529679e-05]

	Penalty	Iteration	Accuracy	Class	Coef	Intercept
20	l2	5	0.710526	[0, 1]	[[-0.00022855554029285456, -0.0004079338711219...	[-2.8391286575467695e-05]
21	l2	10	0.925439	[0, 1]	[[-0.00039096585858844276, -0.00071355747381868...	[-4.891523618198568e-05]
22	l2	20	0.951754	[0, 1]	[[-0.0006880798247416664, -0.00125584069367165...	[-8.716043760648532e-05]
23	l2	50	0.951754	[0, 1]	[[-0.0014149793413827563, -0.00259299643364427...	[-0.00018243325234474316]
24	l2	100	0.951754	[0, 1]	[[-0.002316004072251978, -0.004171963140422094...	[-0.000303311234167857]
25	l2	500	0.956140	[0, 1]	[[-0.0057907084718604365, -0.00883977955075798...	[-0.0007903904337658977]
26	l2	1000	0.964912	[0, 1]	[[-0.007917626428311744, -0.009648200052197587...	[-0.001099362121539854]

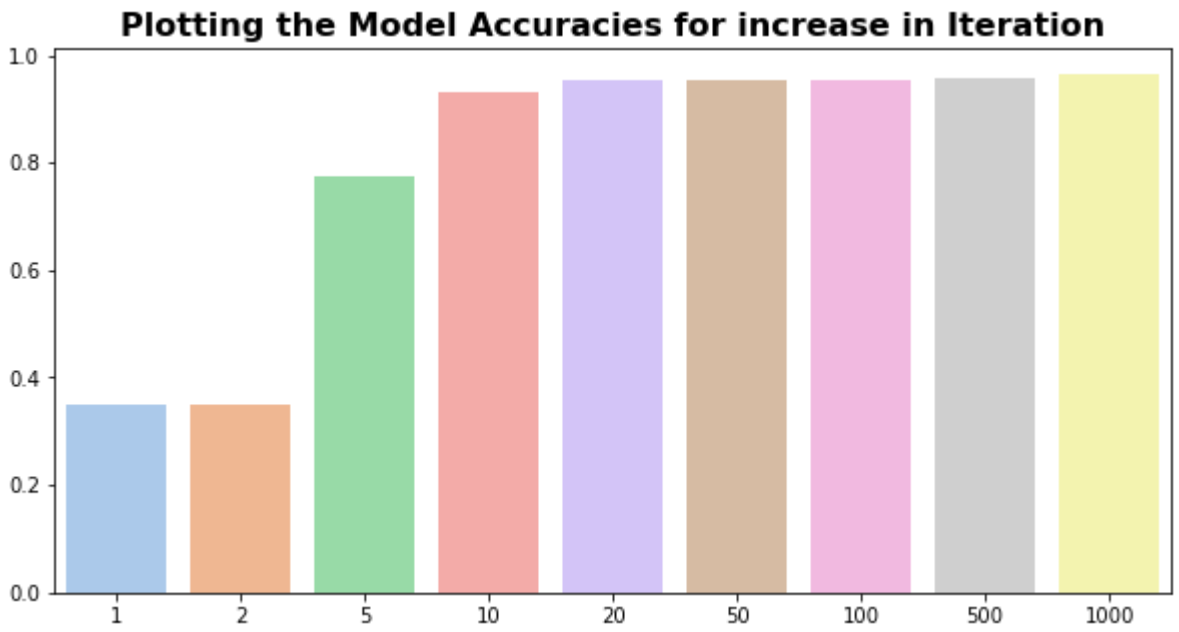
In [19]:

```
Score = []
iterations = [1, 2, 5, 10, 20, 50, 100, 500 , 1000]
for i in range(9):
        log_reg = LogisticRegression(penalty='l2',solver='saga',max_iter=ite
        y_pred = log_reg.predict(X_test)
        Score.append(accuracy_score(y_test, y_pred))
Score
```

```
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn("The max_iter was reached which means "
C:\Users\SRIDHAR\anaconda3\lib\site-packages\sklearn\linear_model\_sag.py:32
8: ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
```

In [20]:
```
plt.figure(figsize = (10,5))
sns.barplot(x = iterations ,y = Score, palette='pastel')
plt.title("Plotting the Model Accuracies for increase in Iteration", fontsiz
```

Out[20]: Text(0.5, 1.0, 'Plotting the Model Accuracies for increase in Iteration')



INFERENCE :

For all the penalty, as the max number of interations increase the accuracy score also increases.

REFERENCES:

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
(https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)
https://www.kaggle.com/code/aditimulye/breast-cancer-prediction
(https://www.kaggle.com/code/aditimulye/breast-cancer-prediction)