Lasso Regression

In [8]:
```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')

from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.linear_model import Lasso

data = pd.read_csv('teams.csv')
X = data.iloc[:,0:8]
Y = data['medals']
```

In [9]:
```
X
```

Out[9]:

	team	year	athletes	events	age	height	weight	prev_medals
0	AFG	1964	8	8	22.0	161.0	64.2	0.0
1	AFG	1968	5	5	23.2	170.2	70.0	0.0
2	AFG	1972	8	8	29.0	168.3	63.8	0.0
3	AFG	1980	11	11	23.6	168.4	63.2	0.0
4	AFG	2004	5	5	18.6	170.8	64.8	0.0
...	...	...	...	...	...	...	...	...
2009	ZIM	2000	26	19	25.0	179.0	71.1	0.0
2010	ZIM	2004	14	11	25.1	177.8	70.5	0.0
2011	ZIM	2008	16	15	26.1	171.9	63.7	3.0
2012	ZIM	2012	9	8	27.3	174.4	65.2	4.0
2013	ZIM	2016	31	13	27.5	167.8	62.2	0.0

2014 rows × 8 columns

In [10]:
```
Y
```

Out[10]:
```
0       0
1       0
2       0
3       0
4       0
       ..
2009    0
2010    3
2011    4
2012    0
2013    0
Name: medals, Length: 2014, dtype: int64
```

In [4]:
```
lasso_model = Lasso(alpha=0.02)

cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)

scores = cross_val_score(lasso_model, X, Y, scoring='neg_mean_absolute_error', cv=cv, n_
```

```
scores = np.absolute(scores)
print('Mean MAE: %.3f (%.3f)' % (np.mean(scores), np.std(scores)))
```

```
Mean MAE: 3.380 (0.521)
```

Ridge regression

In [5]:
```
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.linear_model import Ridge
import pandas as pd
import numpy as np

data = pd.read_csv('teams.csv')
X = data.iloc[:,0:8]
Y = data['medals']

ridge_model = Ridge(alpha=0.02)

cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)

scores = cross_val_score(ridge_model, X, Y, scoring='neg_mean_absolute_error', cv=cv, n_

scores = np.absolute(scores)
print('Mean MAE: %.3f (%.3f)' % (np.mean(scores), np.std(scores)))
```

```
Mean MAE: 3.390 (0.529)
```

Linear Regression

In [6]:
```
from sklearn.linear_model import Lasso
from sklearn.linear_model import LinearRegression

data = pd.read_csv('teams.csv')
X = data.iloc[:,0:8]
Y = data['medals']

lr_model = LinearRegression()

cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)

scores = cross_val_score(lr_model, X, Y, scoring='neg_mean_absolute_error', cv=cv, n_job

scores = np.absolute(scores)
print('Mean MAE: %.3f (%.3f)' % (np.mean(scores), np.std(scores)))
```

```
Mean MAE: 3.391 (0.530)
```

Analysis: We can see that from above results that the lasso regression has best output among other regressions