In [26]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from sklearn.metrics import r2_score
```

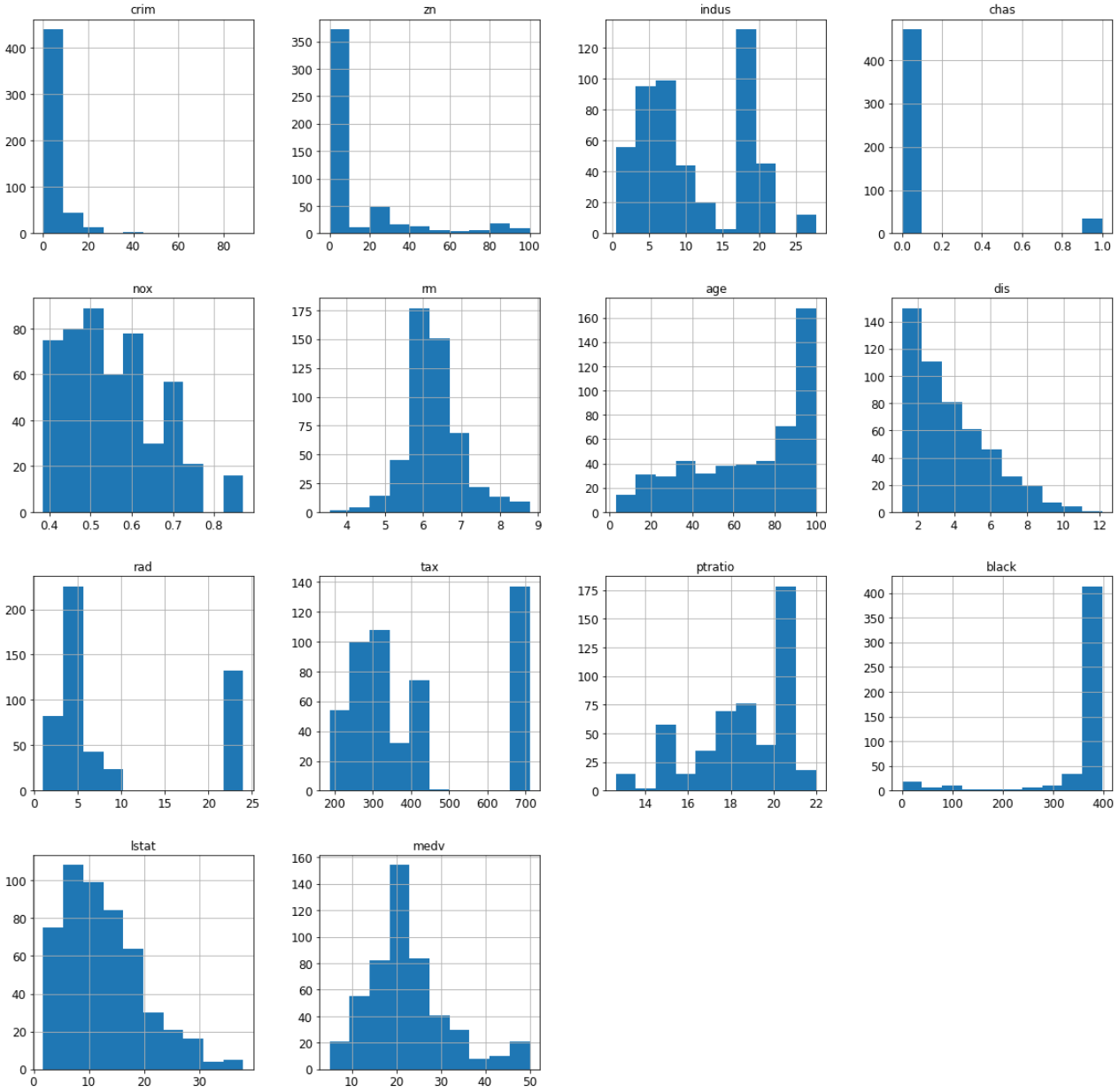
In [27]:
```
data = pd.read_csv('Boston.csv')
data = data.drop(['Unnamed: 0'],axis=1)
data.head(5)
```

Out[27]:

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2

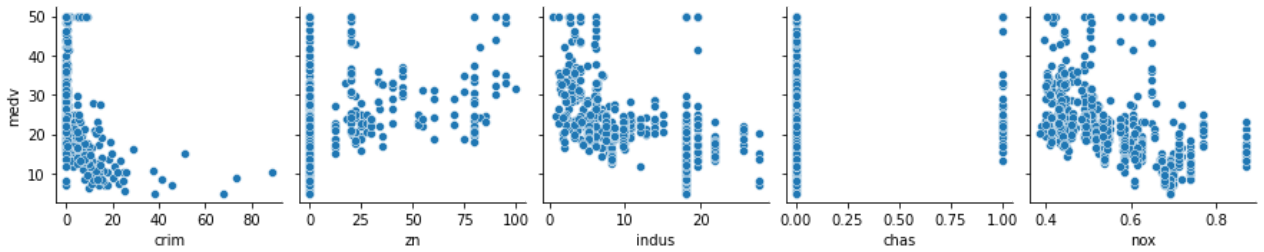
In [28]:
```
data.hist(figsize = (20,20),xlabelsize=12, ylabelsize=12)
```

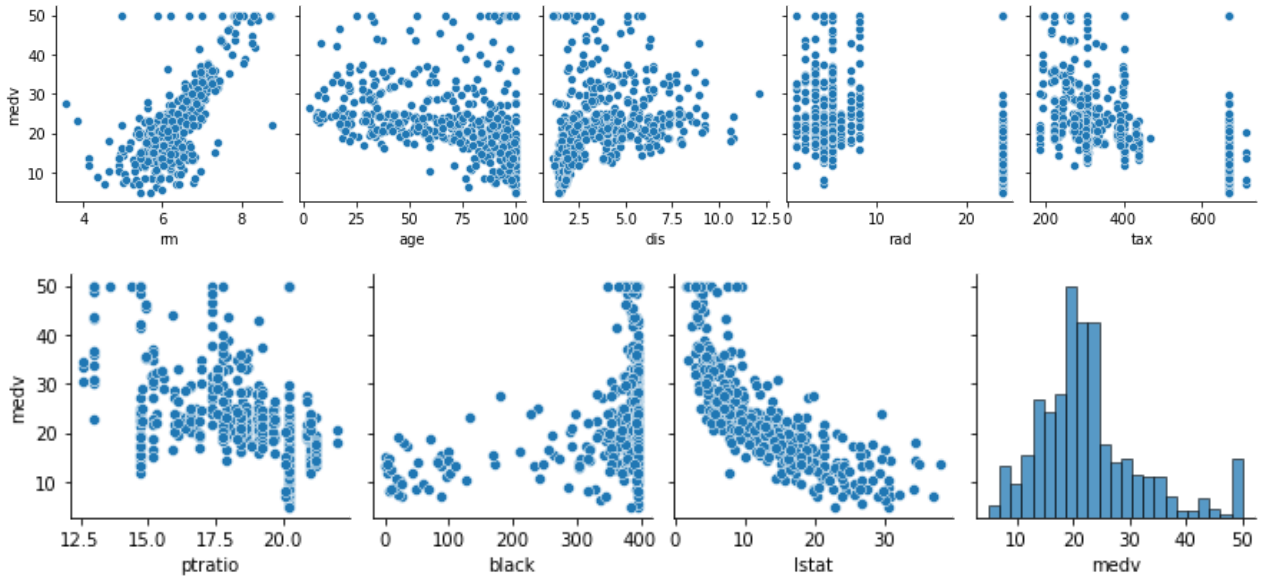
Out[28]:
```
array([[<AxesSubplot:title={'center':'crim'}>,
        <AxesSubplot:title={'center':'zn'}>,
        <AxesSubplot:title={'center':'indus'}>,
        <AxesSubplot:title={'center':'chas'}>],
       [<AxesSubplot:title={'center':'nox'}>,
        <AxesSubplot:title={'center':'rm'}>,
        <AxesSubplot:title={'center':'age'}>,
        <AxesSubplot:title={'center':'dis'}>],
       [<AxesSubplot:title={'center':'rad'}>,
        <AxesSubplot:title={'center':'tax'}>,
        <AxesSubplot:title={'center':'ptratio'}>,
        <AxesSubplot:title={'center':'black'}>],
       [<AxesSubplot:title={'center':'lstat'}>,
        <AxesSubplot:title={'center':'medv'}>, <AxesSubplot:>,
        <AxesSubplot:>]], dtype=object)
```



In [29]:

```
for i in range(0, len(data.columns), 5):
    sns.pairplot(data=data,
                 x_vars=data.columns[i:i+5],
                 y_vars=['medv'])
```

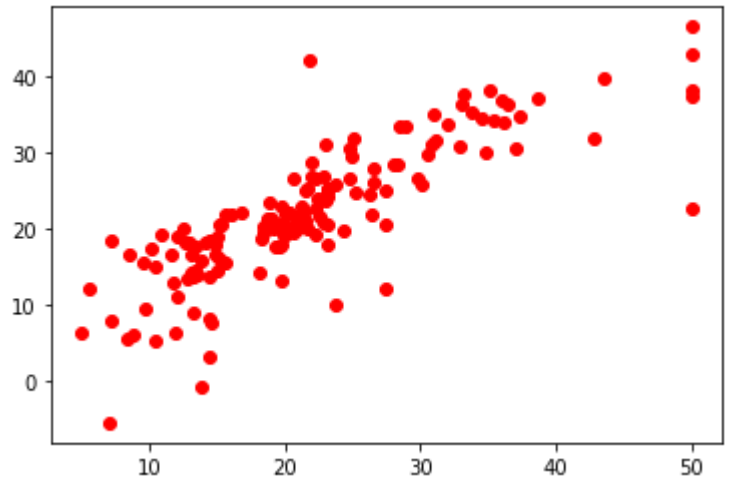




In [30]:
```
y=data['medv']
data = data.drop('medv',axis = 1)
```

In [31]:
```
x_train,x_test,y_train,y_test=train_test_split(data,y,test_size=0.3)
reg=LinearRegression()
reg.fit(x_train,y_train)
y_pred = reg.predict(x_test)
plt.scatter(y_test, y_pred, c = 'red')
```

Out[31]: <matplotlib.collections.PathCollection at 0x21be858e710>



In [32]:
```
print('MAE', metrics.mean_absolute_error(y_test, y_pred))
print('MSE', metrics.mean_squared_error(y_test, y_pred))
print('RMSE', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
print('R squared error', r2_score(y_test, y_pred))
```

```
MAE 3.7129656398788566
MSE 29.390148626967523
RMSE 5.4212681751567615
R squared error 0.66708551054697
```