In [1]:

```
import numpy as np
import pandas as pd
import datetime
from sklearn.cross_validation import KFold
from sklearn.cross_validation import train_test_split
import time
from sklearn import preprocessing
from scipy.stats import skew
```

In [2]:

```
train = pd.read_csv("./input/train.csv") # read train data
test = pd.read_csv("./input/test.csv") # read test data

tables = [train,test]
print ("Delete features with high number of missing values...")
total_missing = train.isnull().sum()
to_delete = total_missing[total_missing>(train.shape[0]/3.)]
for table in tables:
    table.drop(list(to_delete.index),axis=1,inplace=True)

numerical_features = test.select_dtypes(include=["float","int","bool"]).column
s.values
categorical_features = train.select_dtypes(include=["object"]).columns.values
```

```
Delete features with high number of missing values...
```

In [3]:

```
to_delete
```

Out[3]:

```
Alley          1369
FireplaceQu     690
PoolQC         1453
Fence          1179
MiscFeature    1406
dtype: int64
```