In [4]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [9]:

```
df = pd.read_csv("MNIST_test.csv")
```

In [10]:

```
df.columns
```

Out[10]:

```
Index(['pixel0', 'pixel1', 'pixel2', 'pixel3', 'pixel4', 'pixel5', 'pixel6',
       'pixel7', 'pixel8', 'pixel9',
       ...
       'pixel774', 'pixel775', 'pixel776', 'pixel777', 'pixel778', 'pixel779',
       'pixel780', 'pixel781', 'pixel782', 'pixel783'],
      dtype='object', length=784)
```

In [28]:

```
#Checking the shape and having a look at data.
print(df.head())
print(df.shape)
```

```
   pixel0  pixel1  pixel2  pixel3  pixel4  pixel5  pixel6  pixel7  pixel8  \
0       0       0       0       0       0       0       0       0       0
1       0       0       0       0       0       0       0       0       0
2       0       0       0       0       0       0       0       0       0
3       0       0       0       0       0       0       0       0       0
4       0       0       0       0       0       0       0       0       0

   pixel9    ...     pixel774  pixel775  pixel776  pixel777  pixel778  \
0       0    ...            0         0         0         0         0
1       0    ...            0         0         0         0         0
2       0    ...            0         0         0         0         0
3       0    ...            0         0         0         0         0
4       0    ...            0         0         0         0         0

   pixel779  pixel780  pixel781  pixel782  pixel783
0         0         0         0         0         0
1         0         0         0         0         0
2         0         0         0         0         0
3         0         0         0         0         0
4         0         0         0         0         0

[5 rows x 784 columns]
(28000, 784)
```

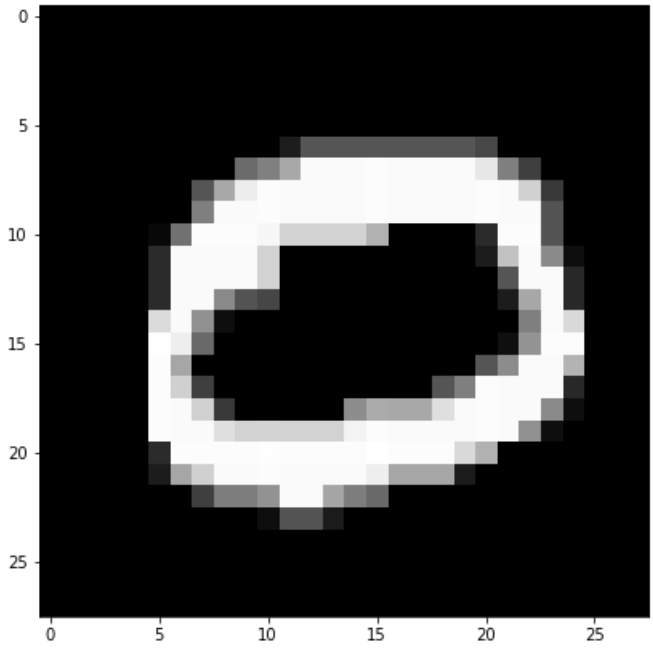
In [26]:

```
#Plotting one entry of the data.

plt.figure(figsize=(7,7))
idx = 1

new = df.iloc[idx].as_matrix().reshape(28,28)
plt.imshow(new,cmap='gray',interpolation='none')
```

```
C:\Users\omkar reddy\Anaconda3-1\lib\site-packages\ipykernel_launcher.py:4: FutureWarning: Method
.as_matrix will be removed in a future version. Use .values instead.
  after removing the cwd from sys.path.
```

Out[26]:

```
<matplotlib.image.AxesImage at 0x24d0045bc18>
```



In [33]:

```
#Standardization of data or Data Preprocessing

from sklearn.preprocessing import StandardScaler
std_df = StandardScaler().fit_transform(df)
```

In [34]:

```
std_df.shape
```

Out[34]:

```
(28000, 784)
```

In [37]:

```
# Performing PCA for dimensionality reduction
#Let's reduce 784 dimensions to 10

from sklearn import decomposition
p= decomposition.PCA()
p.n_components = 10
p_df = p.fit_transform(std_df)

print("reduced df shape ", p_df.shape)
```

```
reduced df shape  (28000, 10)
```

As you can see the the 784 dimensions are reduced to 10 with the help of PCA.

Note: The above exercise is to show the implementation of PCA, the resultant 10 dimensions are not the best dimensions for the model.