In [40]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn import linear_model
```

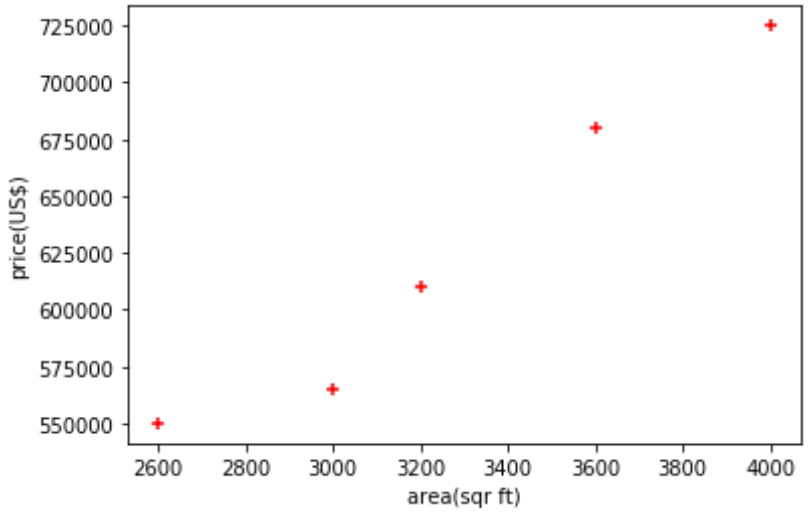
In [41]:
```
df = pd.read_csv("C:/Users/prasa/Desktop/ds projects/panda/homeprices.csv")
df
```

Out[41]:

	area	price
0	2600	550000
1	3000	565000
2	3200	610000
3	3600	680000
4	4000	725000

In [42]:
```
plt.xlabel('area(sqr ft)')
plt.ylabel('price(US$)')
plt.scatter(df.area,df.price,color='red',marker='+')
```

Out[42]:
```
<matplotlib.collections.PathCollection at 0x25caf529388>
```



In [43]:
```
reg = linear_model.LinearRegression()
reg.fit(df[['area']],df.price)
```

Out[43]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normaliz
e=False)
```

In [44]:
```
reg.predict([[5000]])
```

Out[44]:
```
array([859554.79452055])
```

In [45]:
```
reg.coef_ #value of m slope
```

Out[45]:
```
array([135.78767123])
```

In [46]:
```
reg.intercept_ # b intercept
```

Out[46]:
```
180616.43835616432
```

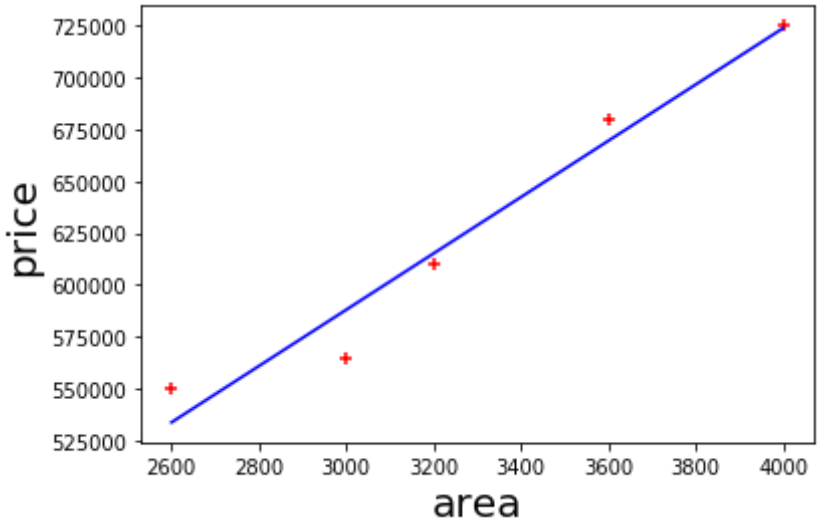
Equation for price is "price= m*area+b" ("y = mx + b")

In [47]:
```
135.78767123*5000+180616.43835616432
```

Out[47]:
```
859554.7945061643
```

In [48]:
```
plt.xlabel('area',fontsize=20)
plt.ylabel('price',fontsize=20)
plt.scatter(df.area,df.price,color='red',marker='+')
plt.plot(df.area,reg.predict(df[['area']]),color='blue')
```

Out[48]:
```
[<matplotlib.lines.Line2D at 0x25caf575e08>]
```



In [29]:
```
d= pd.read_csv("C:/Users/prasa/Desktop/ds projects/panda/areas.csv")
d.head(3)
```

Out[29]:

	area
0	1000
1	1500
2	2300

In [32]:
```
p=reg.predict(d)
```

In [34]:
```
d['prices'] = p
```

In [35]:
```
d
```

Out[35]:

	area	prices
0	1000	3.164041e+05
1	1500	3.842979e+05
2	2300	4.929281e+05
3	3540	6.613048e+05
4	4120	7.400616e+05
5	4560	7.998082e+05
6	5490	9.260908e+05
7	3460	6.504418e+05
8	4750	8.256079e+05
9	2300	4.929281e+05
10	9000	1.402705e+06
11	8600	1.348390e+06
12	7100	1.144709e+06

In [37]:
```
d.to_csv("C:/Users/prasa/Desktop/ds projects/panda/areas.csv")
```