In [10]:

```
%matplotlib notebook
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split

fruits=pd.read_table('fruit_data_with_colors.txt')
```

In [11]:

```
from matplotlib import cm
#feature
X=fruits[['mass','width','height','color_score']]
Y=fruits['fruit_label']
X_train, X_test, Y_train, Y_test=train_test_split(X,Y,random_state=0)
```

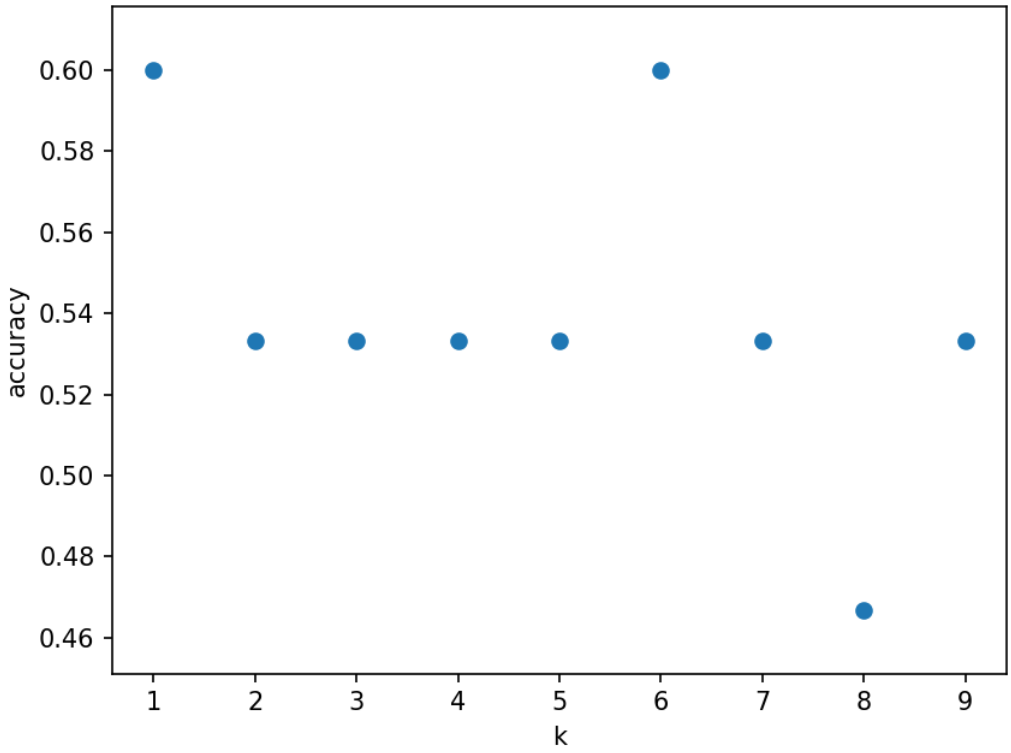
In [24]:

```
#parameter change
k_range=range(1,10)
scores=[]

for k in k_range:
    knn=KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train,Y_train)
    scores.append(knn.score(X_test,Y_test))
```

In [25]:

```
plt.figure()
plt.xlabel('k')
plt.ylabel('accuracy')
plt.scatter(k_range,scores)
```



Out[25]:

```
<matplotlib.collections.PathCollection at 0x23a9f5c5cc8>
```

In [ ]:

```
#k가 1,6일때 가장 좋은 성능을 보입니다.
```

In [26]:

```
scores
```

Out[26]:

```
[0.6,
 0.5333333333333333,
 0.5333333333333333,
 0.5333333333333333,
 0.5333333333333333,
 0.6,
 0.5333333333333333,
 0.4666666666666667,
 0.5333333333333333]
```

In [28]:

```
lookup_fruit_name=dict(zip(fruits.fruit_label.unique(), fruits.fruit_name.unique()))
fruit_prediction=knn.predict([[120,6,8,0.7]])
lookup_fruit_name[fruit_prediction[0]]
```

Out[28]:

```
'lemon'
```

In [ ]:

```

```