

Question 1

In [1]:

```
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

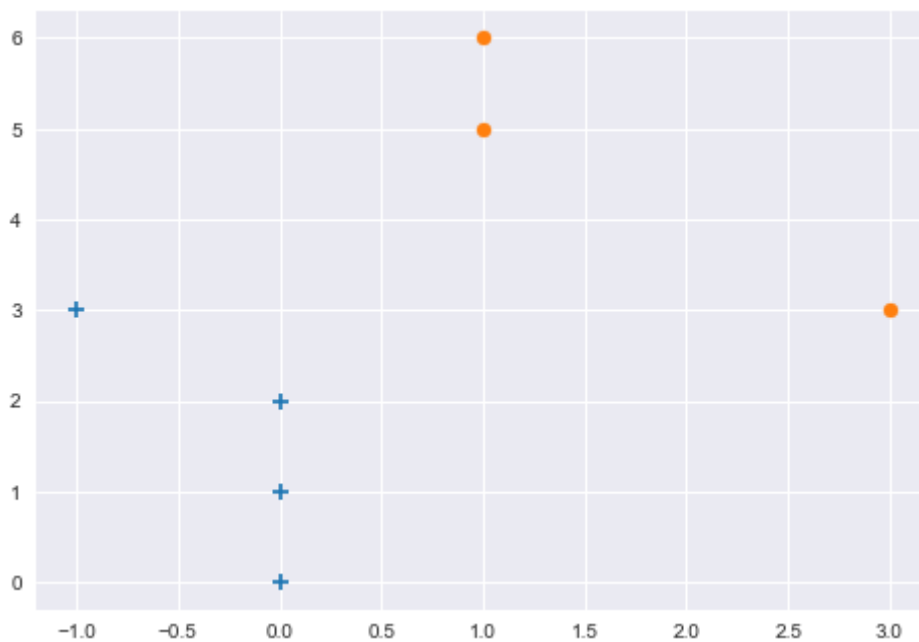
plt.style.use('seaborn')
```

In [2]:

```
Xp = np.array([[-1, 3], [0, 2], [0, 1], [0, 0]])
Xn = np.array([[1, 5], [1, 6], [3, 3]])
```

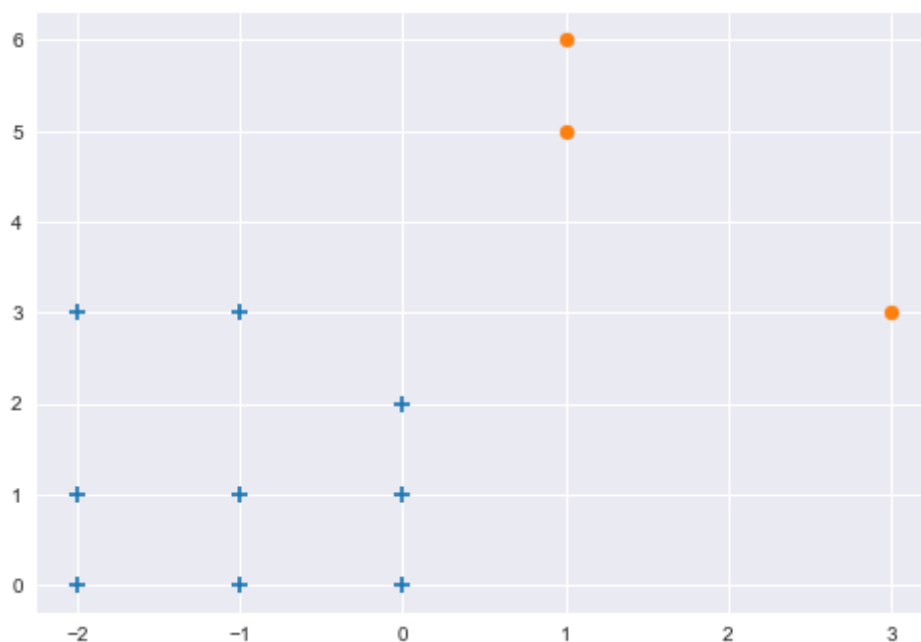
In [3]:

```
plt.scatter(Xp[:, 0], Xp[:, 1], marker='+', c='#1f77b4', s=50);
plt.scatter(Xn[:, 0], Xn[:, 1], marker='o', c='#ff7f0e', s=50);
```



In [4]:

```
Xp = np.array([[-1, 3], [0, 2], [0, 1], [0, 0], [-2, 0], [-2, 1], [-2, 3], [-1, 0], [-1, 1]])
plt.scatter(Xp[:, 0], Xp[:, 1], marker='+', c='#1f77b4', s=50);
plt.scatter(Xn[:, 0], Xn[:, 1], marker='o', c='#ff7f0e', s=50);
```



Question 3

In [7]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC # "Support Vector Classifier"
```

In [8]:

```
df = pd.read_csv('wdbc.csv', header = None)
X = df.iloc[:, 2:32]
Y = df.iloc[:, 1]
```

In [9]:

```
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score

clf1 = SVC(kernel='linear', gamma='auto')
acc_train_sum = 0.0
acc_test_sum = 0.0
pre_sum = 0.0
rec_sum = 0.0
for i in range(0,20):
    (X_train,X_test,Y_train,Y_test)=train_test_split(X, Y, test_size=0.3)
    clf1.fit(X_train, Y_train)
    y_pred = clf1.predict(X_test)

    acc_train_sum += clf1.score(X_train, Y_train)
    acc_test_sum += clf1.score(X_test, Y_test)
    pre_sum += precision_score(Y_test, y_pred, labels=['M','B'], pos_label='M', average='binary',
, sample_weight=None)
    rec_sum += recall_score(Y_test, y_pred, labels=['M','B'], pos_label='M', average='binary', s
ample_weight=None)
#    print(i, "time the score is", clf1.score(X_test, Y_test))
print("Average Accuray train is", acc_train_sum/20)
print("Average Accuray test is", acc_test_sum/20)
print("Average Precision is", pre_sum/20)
print("Average Recall is", rec_sum/20)
```

Average Accuray train is 0.9679648241206034

Average Accuray test is 0.9514619883040932

Average Precision is 0.9617527987603658

Average Recall is 0.9114987073952223

In [10]:

```
map1 = {'M':0.0, 'B':1.0}
Y = Y.map(map1)
(X_train,X_test,Y_train,Y_test)=train_test_split(X, Y, test_size=0.3)
```

In [11]:

```

acc_train_sum = 0.0
acc_test_sum = 0.0
pre_sum = 0.0
rec_sum = 0.0
clf2 = SVC(kernel='rbf', gamma='scale')
for i in range(0,20):
    (X_train,X_test,Y_train,Y_test)=train_test_split(X, Y, test_size=0.3)
    clf2.fit(X_train, Y_train)
    y_pred = clf2.predict(X_test)

    acc_train_sum += clf2.score(X_train, Y_train)
    acc_test_sum += clf2.score(X_test, Y_test)
    pre_sum += precision_score(Y_test, y_pred, labels=[0.,1.], average='micro', sample_weight=None)
    rec_sum += recall_score(Y_test, y_pred, labels=[0.,1.], average='micro', sample_weight=None)
#     print(i, "time the score is", clf1.score(X_test, Y_test))
print("Average Accuray train is", acc_train_sum/20)
print("Average Accuray test is", acc_test_sum/20)
print("Average Precision is", pre_sum/20)
print("Average Recall is", rec_sum/20)

```

Average Accuray train is 0.9149497487437184

Average Accuray test is 0.9116959064327487

Average Precision is 0.9116959064327487

Average Recall is 0.9116959064327487

In [12]:

```
print(y_pred)
```

```

[1. 1. 1. 1. 0. 1. 1. 0. 1. 1. 0. 0. 1. 1. 0. 1. 1. 1. 1. 0. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 0. 1. 1. 0. 0. 0. 1. 1. 0. 1. 1. 1. 0. 1. 0. 1.
 1. 1. 1. 1. 1. 1. 0. 1. 0. 1. 0. 1. 0. 1. 0. 1. 1. 1. 1. 0. 1. 1. 1. 1.
 0. 1. 1. 1. 0. 0. 1. 1. 1. 1. 0. 0. 1. 1. 1. 0. 0. 1. 1. 0. 0. 0. 1. 1.
 1. 0. 0. 0. 0. 1. 1. 0. 1. 1. 0. 1. 1. 1. 1. 1. 0. 1. 0. 1. 0. 1. 1.
 1. 0. 0. 1. 1. 0. 1. 1. 1. 1. 1. 0. 1. 1. 1. 1. 0. 1. 1. 1. 0. 1. 1. 0.
 1. 0. 0. 1. 1. 1. 1. 0. 1. 1. 1. 1. 0. 1. 0. 1. 1. 0. 0. 1. 1. 0. 0. 0.
 1. 1. 1.]

```

In [17]:

```
acc_train_sum = 0.0
acc_test_sum = 0.0
pre_sum = 0.0
rec_sum = 0.0
clf3 = SVC(kernel='rbf', gamma='scale', C=1000)
for i in range(0, 20):
    (X_train, X_test, Y_train, Y_test)=train_test_split(X, Y, test_size=0.3)
    clf3.fit(X_train, Y_train)
    y_pred = clf3.predict(X_test)

    acc_train_sum += clf3.score(X_train, Y_train)
    acc_test_sum += clf3.score(X_test, Y_test)
    pre_sum += precision_score(Y_test, y_pred, labels=[0., 1.], average='micro', sample_weight=None)
    rec_sum += recall_score(Y_test, y_pred, labels=[0., 1.], average='micro', sample_weight=None)
# print(i, "time the score is", clf1.score(X_test, Y_test))
print("Average Accuray train is", acc_train_sum/20)
print("Average Accuray test is", acc_test_sum/20)
print("Average Precision is", pre_sum/20)
print("Average Recall is", rec_sum/20)
```

Average Accuray train is 0.9599246231155778

Average Accuray test is 0.9464912280701754

Average Precision is 0.9464912280701754

Average Recall is 0.9464912280701754

In []: