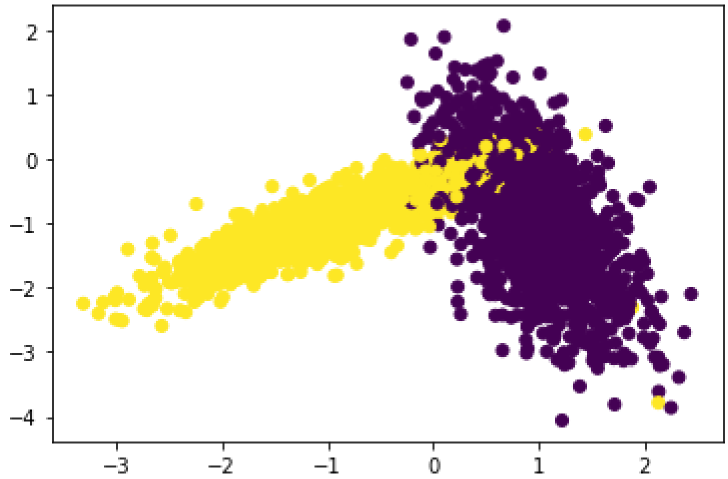
In [1]:

```
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import numpy
from tqdm import tqdm
import numpy as np
from sklearn.metrics.pairwise import euclidean_distances


x,y = make_classification(n_samples=10000, n_features=2, n_informative=2, n_redundant= 0, r
x_train, x_test, y_train, y_test = train_test_split(x,y,stratify=y,random_state=42)

# del X_train,X_test
```

In [2]:

```
%matplotlib inline
import matplotlib.pyplot as plt
colors = {0:'red', 1:'blue'}
plt.scatter(x_test[:,0], x_test[:,1],c=y_test)
plt.show()
```



Implementing Custom RandomSearchCV

```
def RandomSearchCV(x_train,y_train,classifier, param_range, folds):
    # x_train: its numpy array of shape, (n,d)
    # y_train: its numpy array of shape, (n,) or (n,1)
    # classifier: its typically KNeighborsClassifier()
    # param_range: its a tuple like (a,b) a < b
    # folds: an integer, represents number of folds we need to devide the data and
test our model


    #1.generate 10 unique values(uniform random distribution) in the given range
"param_range" and store them as "params"
    # ex: if param_range = (1, 50), we need to generate 10 random numbers in range
```

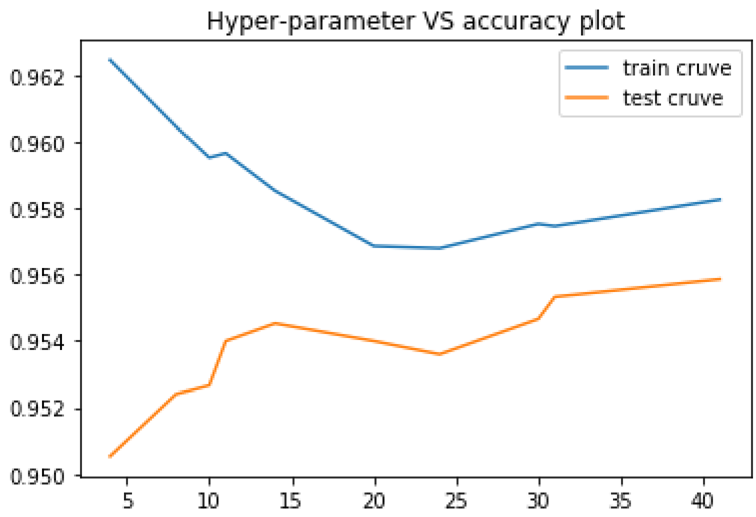
```
1 to 50
    #2.devide numbers ranging from  0 to len(X_train) into groups= folds
    # ex: folds=3, and len(x_train)=100, we can devide numbers from 0 to 100 into
3 groups
      group 1: 0-33, group 2:34-66, group 3: 67-100
    #3.for each hyperparameter that we generated in step 1:
        # and using the above groups we have created in step 2 you will do cross-v
alidation as follows

        # first we will keep group 1+group 2 i.e. 0-66 as train data and group 3:
67-100 as test data, and find train and
          test accuracies

        # second we will keep group 1+group 3 i.e. 0-33, 67-100 as train data and
group 2: 34-66 as test data, and find
          train and test accuracies

        # third we will keep group 2+group 3 i.e. 34-100 as train data and group
1: 0-33 as test data, and find train and
          test accuracies
        # based on the 'folds' value we will do the same procedure

        # find the mean of train accuracies of above 3 steps and store in a list
"train_scores"
        # find the mean of test accuracies of above 3 steps and store in a list "t
est_scores"
    #4. return both "train_scores" and "test_scores"

#5. call function RandomSearchCV(x_train,y_train,classifier, param_range, folds) a
nd store the returned values into "train_score", and "cv_scores"
#6. plot hyper-parameter vs accuracy plot as shown in reference notebook and choos
e the best hyperparameter
#7. plot the decision boundaries for the model initialized with the best hyperpara
meter, as shown in the last cell of reference notebook
```

In [8]:

```
def RandomSearchCV(x_train,y_train, classifier,params, folds):
    train_scores = []
    test_scores = []
    x_train_split = []
    y_train_split = []


    #dividing x_train into groups  :https://stackoverflow.com/questions/1624883/alternative
    for i in range(0, len(x_train), int(len(x_train)/folds)):
        x_train_split.append(x_train[i:i+int(len(x_train)/folds)])
        y_train_split.append(y_train[i:i+int(len(y_train)/folds)])

    # 3.for each hyperparameter that we generated in step 1 and dividing dataset into train

    for parameter in params:
        trainscores_folds = []
        testscores_folds  = []

        for group in range(len(x_train_split)):
            x_train_group = np.concatenate(x_train_split[0:group] + x_train_split[group+1:]
            x_cv_group = x_train_split[group]
            y_train_group = np.concatenate(y_train_split[0:group] + y_train_split[group+1:]
            y_cv_group = y_train_split[group]

            # classifier (K-NN)
            classifier.n_neighbors = parameter
            classifier.fit(x_train_group, y_train_group)

            # Predicton
            Y_pred = classifier.predict(x_cv_group)
            testscores_folds.append(accuracy_score(y_cv_group, Y_pred))

            Y_pred = classifier.predict(x_train_group)
            trainscores_folds.append(accuracy_score(y_train_group, Y_pred))


        train_scores.append(np.mean(np.array(trainscores_folds)))
        test_scores.append(np.mean(np.array(testscores_folds)))

    return train_scores, test_scores
```

In [9]:

```
from sklearn.metrics import accuracy_score
from sklearn.neighbors import KNeighborsClassifier
import matplotlib.pyplot as plt
import random
import warnings
warnings.filterwarnings("ignore")


neigh = KNeighborsClassifier()
params = {'n_neighbors' : tuple(random.sample(range(0, 50), 10))}  # Declaring the paramete
sort_param= tuple(sorted(params['n_neighbors']))
folds = 3

trainscores,testscores = RandomSearchCV(x_train, y_train, neigh,sort_param, folds)



plt.plot(sort_param,trainscores, label='train cruve')
plt.plot(sort_param,testscores, label='test cruve')
plt.title('Hyper-parameter VS accuracy plot')
plt.legend()
plt.show()
```

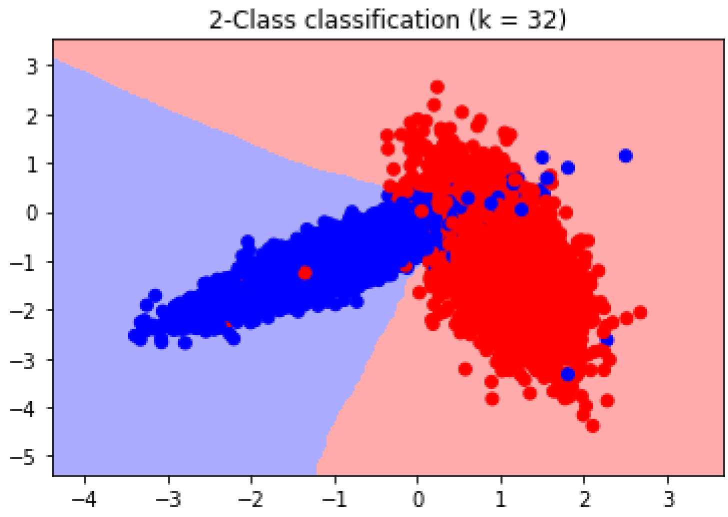


In [10]:

```
# understanding this code line by line is not that importent
def plot_decision_boundary(X1, X2, y, clf):
        # Create color maps
    cmap_light = ListedColormap(['#FFAAAA', '#AAFFAA', '#AAAAFF'])
    cmap_bold = ListedColormap(['#FF0000', '#00FF00', '#0000FF'])

    x_min, x_max = X1.min() - 1, X1.max() + 1
    y_min, y_max = X2.min() - 1, X2.max() + 1

    xx, yy = np.meshgrid(np.arange(x_min, x_max, 0.02), np.arange(y_min, y_max, 0.02))
    Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])
    Z = Z.reshape(xx.shape)

    plt.figure()
    plt.pcolormesh(xx, yy, Z, cmap=cmap_light)
    # Plot also the training points
    plt.scatter(X1, X2, c=y, cmap=cmap_bold)

    plt.xlim(xx.min(), xx.max())
    plt.ylim(yy.min(), yy.max())
    plt.title("2-Class classification (k = %i)" % (clf.n_neighbors))
    plt.show()
```

In [11]:

```
from matplotlib.colors import ListedColormap
neigh = KNeighborsClassifier(n_neighbors = 32)
neigh.fit(x_train, y_train)
plot_decision_boundary(x_train[:, 0], x_train[:, 1], y_train, neigh)
```



In [ ]:

```
```