鸢尾花任意选择3个特征进行特征工程的构建

如下是进行3特征预测的操作流程：

1、导入需要用到的包

|  |
| --- |
| #1、import需要用到的package  from \_\_future\_\_ import print\_function  from sklearn.datasets import load\_iris  from sklearn.neighbors import KNeighborsClassifier  from sklearn.preprocessing import PolynomialFeatures  from sklearn.model\_selection import train\_test\_split  import matplotlib.pyplot as plt |

2、加载模型训练数据

|  |
| --- |
| #加载数据  iris = load\_iris()  X = iris.data  y = iris.target |

3、特征选取

|  |
| --- |
| #特征选取  ploy = PolynomialFeatures(3)  #模型训练  X\_ploy = ploy.fit\_transform(X)  #交叉校验，查找最优的参数  k\_range = range(1,20)  k\_score = []  for k in k\_range:  knn = KNeighborsClassifier(n\_neighbors=k)  scores = cross\_val\_score(knn,X\_ploy,y,cv=10,scoring='accuracy')  k\_score.append(scores.mean())  print(scores)  print(k)  print('-----------')  plt.plot(k\_range,k\_score)  plt.xlabel('Value of k for knn')  plt.ylabel('cross-validated accuracy')  plt.show()  #最优参数是5，进行模型保存  import pickle  knn = KNeighborsClassifier(n\_neighbors=5)  #保存模型  #with open('../resource/atc.pickle','wb') as f:  # pickle.dump(knn,f)  #预测模型  with open('../resource/atc.pickle','rb') as f:  clf2 = pickle.load(f)  print(clf2.predict(X[0:1])) |

如下是输出的曲线：



如图所示找到的最优参数是5。