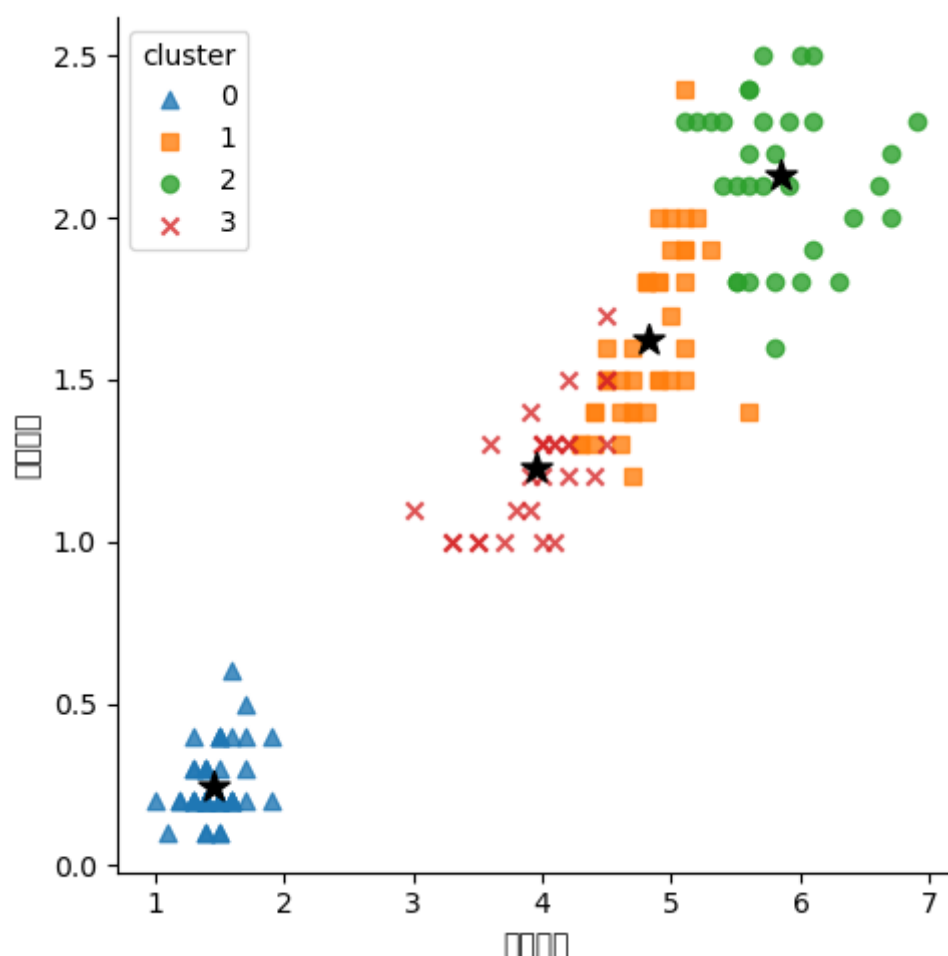


K-mean聚类分析

运行结果



代码解释

这段代码使用Python的pandas, sklearn, matplotlib和seaborn库来进行K-means聚类分析。每个样本有四个特征：花萼长度（sepal length）、花萼宽度（sepal width）、花瓣长度（petal length）和花瓣宽度（petal width）。这些鸢尾花样本分为三种不同的鸢尾花种类。

```
x = iris.drop(labels='Species', axis=1)
```

从iris数据集中删除'Species'列，这一列用于验证结果，将其余特征作为用于聚类建模的数据集X。

```
kmeans = KMeans(n_clusters=4)
kmeans.fit(X)
```

使用KMeans算法构建了一个聚类模型，并对数据集X进行训练。

```
x['cluster'] = kmeans.labels_
```

将聚类模型的结果标签存储在X数据集的'cluster'列中

```
x.cluster.value_counts()
```

统计各个聚类类别的频数，即每个类别中样本的数量。

```
centers = kmeans.cluster_centers_
```

获取聚类模型中每个簇的簇中心

```
sns.lmplot(x='Petal_Length', y='Petal_width', hue='cluster', markers=['^', 's',  
                                'o', 'x'],  
           data=X, fit_reg=False, scatter_kws={'alpha': 0.8}, legend_out=False)  
plt.scatter(centers[:, 2], centers[:, 3], marker='*', color='black', s=130)  
plt.xlabel('花瓣长度')  
plt.ylabel('花瓣宽度')  
plt.show()
```

使用seaborn库绘制散点图，其中x轴表示花瓣长度，y轴表示花瓣宽度，不同聚类簇的样本用不同的标记符号表示（'^'、's'、'o'、'x'），而聚类的结果标签由'hue'参数指定。通过散点图可以直观地展示不同簇的分布情况。另外，plt.scatter函数用于在图中标注簇中心。