

19521216 – Trần Huỳnh Kỳ Anh

Phân biệt parameters và hyper parameters:

Các tham số của mô hình được ước tính dựa trên dữ liệu trong quá trình đào tạo mô hình và các siêu tham số của mô hình được đặt theo cách thủ công và được sử dụng trong các quy trình để giúp ước tính các tham số của mô hình.

Hyperparameters Các siêu thông số lấy từ Scikit KMeans:

```
class sklearn.cluster.KMeans(n_clusters=8, init='k-means++', n_init=10, max_iter=300, tol=0.0001, precompute_distances='auto', verbose=0, random_state=None, copy_x=True, n_jobs=None, algorithm='auto')
```

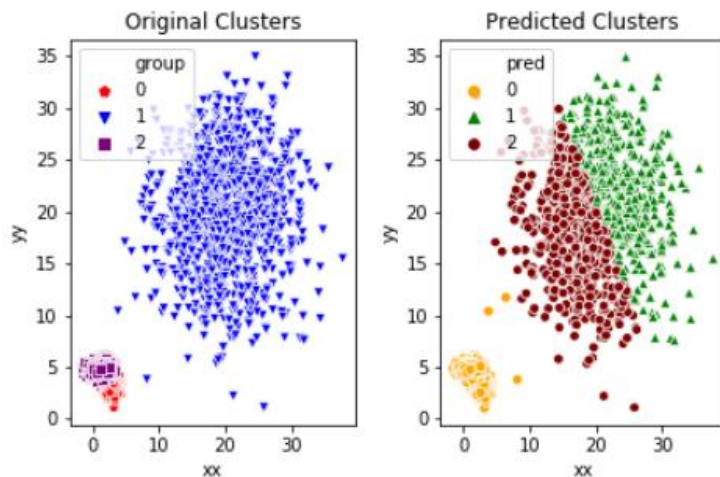
random_state

Đây là thiết lập một hạt giống ngẫu nhiên. Sẽ rất hữu ích nếu chúng ta muốn tái tạo các cụm chính xác nhiều lần.

n_clusters

Chúng ta cần cung cấp cho thuật toán số lượng cụm mà chúng ta muốn. Ví dụ với `n_clusters = 3`:

```
km = KMeans(n_clusters=3, random_state=1234).fit(dftmp.loc[:, dftmp.columns != 'group'])
```



init

Tham số bạn có thể đặt các trọng tâm cụm ban đầu.

Ngoài ra

Có các siêu tham số khác như `tol`, `max_iter` giúp hỗ trợ thời gian tính toán. Những thông số này trở nên quan trọng hơn trong một vấn đề phức tạp hơn.