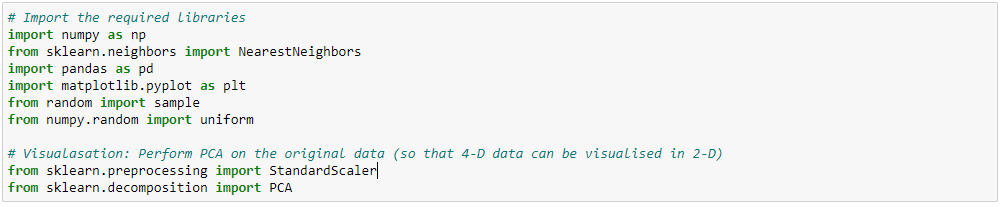
# کتابخانه‌های مورد نیاز

ابتدا کتابخونه‌های لازم توی برنامه را اضافه میکنیم



# روش کار الگوریتم

* یه داده اولیه داریم به اندازه n و به اسم X
* بدون عمل جایگذاری حدود m تعداد از داده اولیه رو که کوچکتر از اندازه کل داده است (m < n) انتخاب میکنیم، در اینجا ما حدود ۵٪ رو انتخاب کردیم.
* به صورت تصادفی یه سری دیتای جدید رو تولید میکنیم و اسمش رو میزاریم Y که به صورت یکنواخت توزیع شدند.
* دو روش فاصله رو تعریف میکنمی  
  - U که فاصله یک نقطه در Y از نزدیک ترین همایسه های اون نقطه در X  
  - W که قاصه یک نقطه از X از نزدیک ترین همسابه های اون نقطه در X
* اگر داده ها d بعدی باشند. **Hopkins statistic به صورت زیر تعریف میشه**



**Hopkins statistic: روشی برای محاسبه و اندازه گیری cluster tendency مجموعه داده ها.**



توضیح الگوریتم بالا داده شد. طبق فرمول داره محاصبه میشه. یه داده X رو میگیره . بعد مشخص میکنه 5% از اندازه داده چقدر میشه، مثلا اگر 100 تا داده داریم، میشه 5 تا، بعد میاد به اون تعداد از داده اصلی جدا میکنه . بعدش هم با استفاده از knn فاصله ها رو حساب میکنه و فرمولی که عکسش بالا اومده رو محاسبه میکنه.

# اطلاعات دیتاست

* 150 تا نمونه گل توی سه تا دسته (برای هر دسته 50 تا)
* 4 تا ویژگی عددی و 1 دونه هم پیشبینی ( کلا 5 تا)
* داده missing value هم جزو ویژگی ها نداریم

# اطلاعات ویژگی ها

* زنبق نوک زبر (Iris setosa)
* زنبق رنگارنگ (Iris versicolor)
* زنبق ویرجینیا (Iris virginica)

# نمایش دیتاست روی نمودار و پلات

# ایجاد داده تصادفی و نمایش روی نمودار و پلات