

## 实验 6:

1. 采用局部加权线性回归, 预测鲍鱼年龄(使用鲍鱼年龄数据集 `abalone.txt`), 随机选取部分数据分别用于训练、余下数据测试。

采用不同核大小(不同  $k$  值), 分别计算训练误差和测试误差。

2. 使用二分 `k-means` 算法对地图上的点聚类

使用 `places.txt` 文件中地图上的点的纬度和经度数据(第 4 列、第 5 列), 采用二分 `k-means` 对地图上的点聚类( $k$  设定为 5)。

3. 采用 `sklearn` 中 `make_moons` 生成半月形数据集, 分别采用 `kmeans` 聚类和 `DBSCAN` 聚类, 采用 `matplotlib` 绘制图形聚类结果, 比较聚类效果。

4. 利用 PCA 对半导体制造数据集 `secom.data`

(<http://archive.ics.uci.edu/ml/machine-learning-databases/secom>)

进行降维。注:该数据包含了较多缺失值, 采用平均值对所有缺失值进行替换, 平均值由非缺失的数据得到。在实验中取不同的主成分截断值来检验性能。采用自编 `python` 代码和使用 `sklearn` 库分别实现。

5.对于用户和菜肴打分的场景可构建如下图所示矩阵,

	Unagi Don	Chicken Katsu	Chirashi	Tri Tip	Salmon Burger	Ruben	Chicken Tandoori	Mapo Tofu	Kung Pao Chicken	Paneer Jalfrazie	Big Dutchman
Brett	2	0	0	4	4	0	0	0	0	0	0
Rob	0	0	0	0	0	0	0	0	0	0	5
Drew	0	0	0	0	0	0	0	1	0	4	0
Scott	3	3	4	0	3	0	0	2	2	0	0
Mary	5	5	5	0	0	0	0	0	0	0	0
Brent	0	0	0	0	0	0	5	0	0	5	0
Kyle	4	0	4	0	0	0	0	0	0	0	5
Sara	0	0	0	0	0	4	0	0	0	0	4
Shaney	0	0	0	0	0	0	5	0	0	5	0
Brendan	0	0	0	3	0	0	0	0	4	5	0
Leanna	1	1	2	1	1	2	1	0	4	5	0

对该矩阵进行 SVD 分解降维，在低维空间下利用 item-based 协同推荐方法进行推荐。（采用余弦相似性进行计算）

6. 使用 Scikit-learn 库中的 Olivetti faces 数据集，选取该数据集中的一张图像，使用 SVD 进行压缩重构，观察和比较不同奇异值数量的压缩效果。