zfmlhw06

1. 阅读预习题

阅读 文章 http://blog.csdn.net/baimafujinji/article/details/50570824 (http://blog.csdn.net/baimafujinji/article/details/50570824) 中前半部分k-means算法计算的实例。
提交方式：提交一张你阅读界面的截图。

原 K-means算法原理与R语言实例

2016年01月23日 17:57:47 白马负金羁 阅读数：33237 标签： K-means n语言 数据挖掘 机器学习

更多

聚类是将相似对象归到同一个簇中的方法，这有点像全自动分类。簇内的对象越相似，聚类的效果越好。支持向量机、神经网络所讨论的分类问题都是有监督的学习方式，现在我们所介绍的聚类则是无监督的。其中，K均值（*K*-means）是最基本、最简单的聚类算法。

学习更多机器学习算法原理并了解在R中如何实现机器学习的技术，你还可以参考我的《R语言实战：机器学习与数据分析》（电子工业出版社出版）一书。

在K均值算法中，质心是定义聚类原型（也就是机器学习获得的结果）的核心。在介绍算法实施的具体过程中，我们将演示质心的计算方法。而且你将看到除了第一次的质心是被指定的以外，此后的质心都是经由计算均值而获得的。

https://blog.csdn.net/baimafujinji/article/details/77876528 (https://blog.csdn.net/baimafujinji/article/details/77876528)

原 机器学习中的k-means聚类及其Python实例

2017年09月07日 07:22:20 白马负金羁 阅读数：3182 标签： k-means 数据挖掘 Python 机器学习 scikit-learn 更多

在2006年12月召开的 IEEE 数据挖掘国际会议上（ICDM， International Conference on Data Mining），与会的各位专家选出了当时的十大“数据挖掘算法”（ top 10 data mining algorithms ）， *k*-means算法即位列其中。

该算法思路简洁，但是在实践中却相当有效。如果你对其算法原理仍不甚了解，你可以参考本博客之前的文章《K-means算法原理与R语言实例》。在此前的文章中，我们给出的实例是基于R语言实现的。本文将演示在Python语言中利用scikit-learn提供的函数来进行基于 *k*-means之机器学习的实例。最后，本文还会演示 *k*-means算法在图像处理中的一个重要应用。

2. 编程实践题（*20%）

对鸢尾花数据集（iris dataset）做 k-means 聚类。
注意事项和要求，不能满足题目要求的提交内容会被黄牌扣罚：
1、聚类是非监督学习，所以导入数据时，不需要带label。
2、请使用 R、Python 或者 MATLAB。
3、请仅仅使用后两个特征进行聚类（即petal.length和petal.width）。
4、绘制图形以可视化地显示你的聚类效果。
5、提交完整的（包括引用必要头文件所需的代码）可以执行的代码，代码部分请不要以截图方式提交，因为无法复制粘贴而不能运行的代码将无法判定正确与否。

In [7]:

```
from sklearn import datasets
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

In [2]:

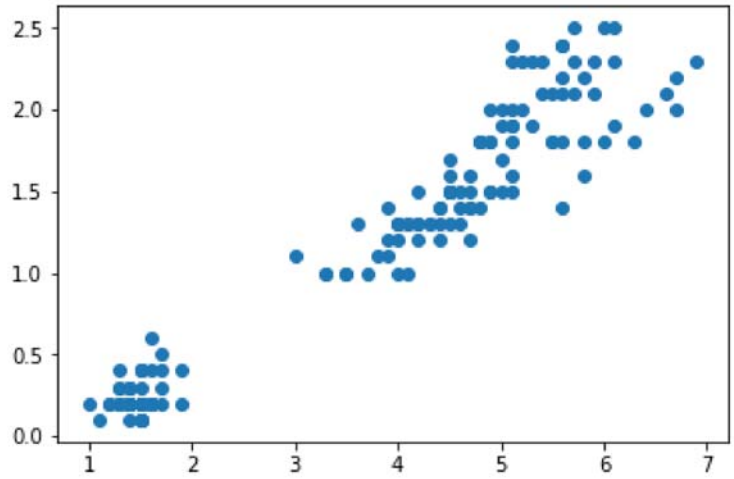
```
iris = datasets.load_iris() # 导入数据集
X = iris.data # 获得其特征向量
y = iris.target # 获得样本labe
```

In [6]:

```
#取后两个特征
X1 = X[:, 2:]
plt.scatter(X1[:,0], X1[:, 1])
```

Out[6]:

```
<matplotlib.collections.PathCollection at 0x7f583b7012b0>
```



In [9]:

```
model = KMeans(n_clusters=3, random_state=0)
y_pred = model.fit_predict(X1)
y_pred
```

Out[9]:

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2,
2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 2, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], dtype=i
nt32)
```

In [10]:

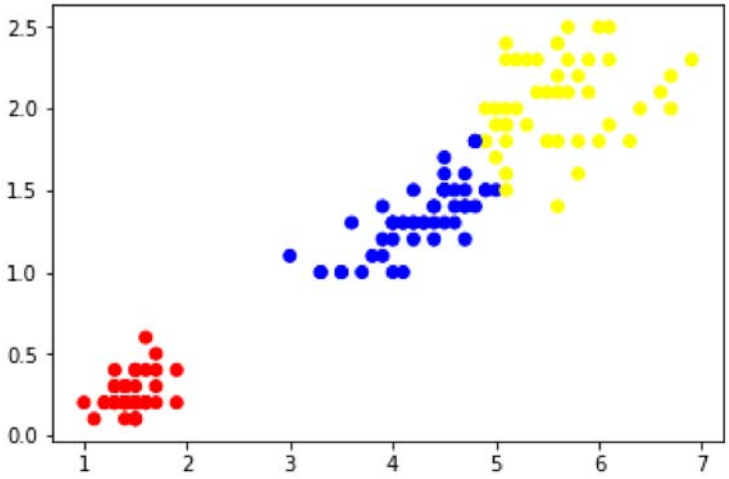
```
color = ('red', 'yellow', 'blue')
colors = [color[i] for i in y_pred]
```

In [13]:

```
plt.scatter(X1[:,0], X1[:,1], color=colors)
```

Out[13]:

```
<matplotlib.collections.PathCollection at 0x7f583690a0f0>
```



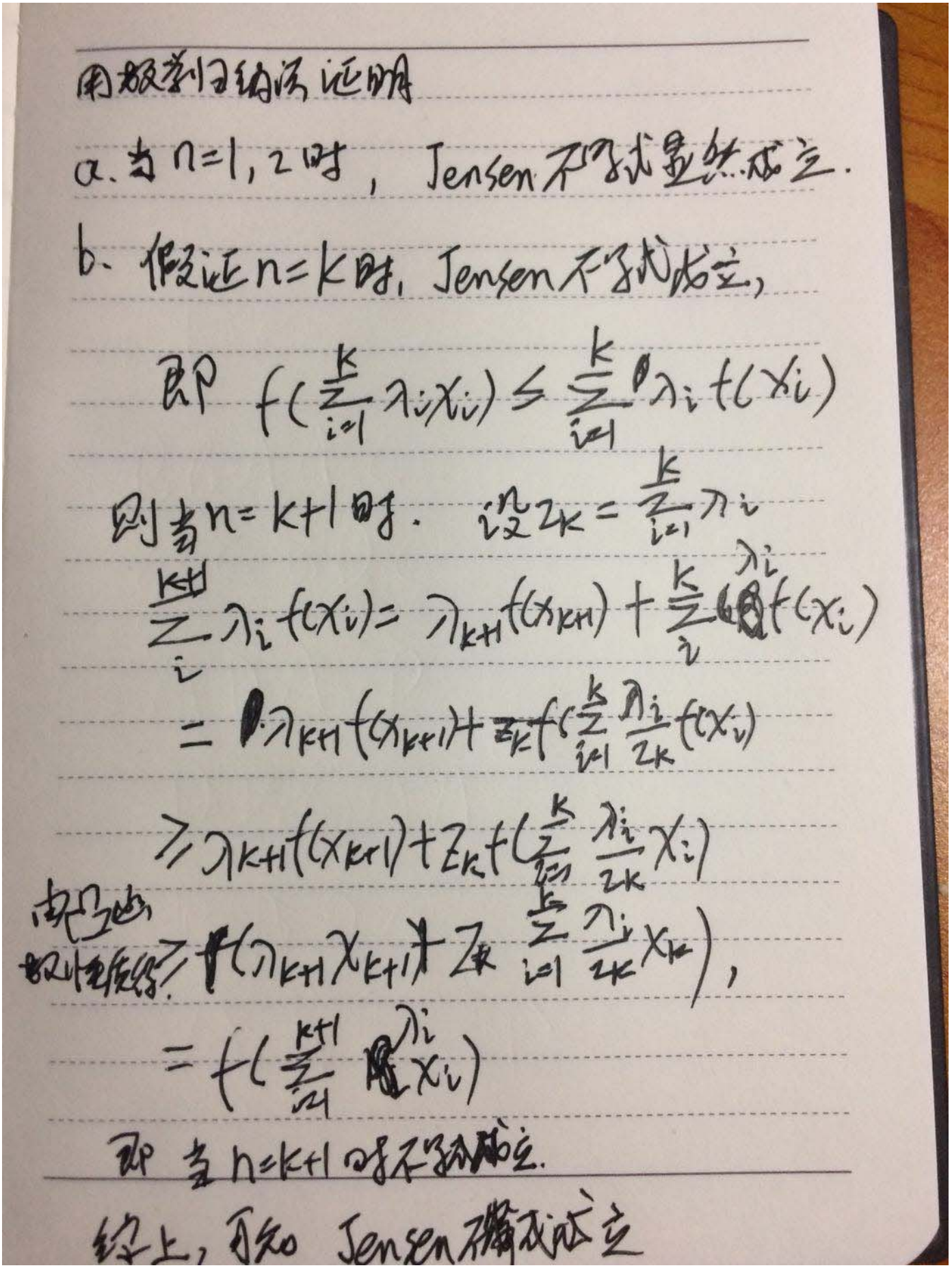
3、证明题（*35%）

证明詹森不等式，该不等式的内容描述如下：

从凸函数的性质中所引申出来的一个重要结论就是詹森(Jensen)不等式：如果 f 是定义在实数区间$[a,b]$上的连续凸函数，$x_1,x_2,\cdots,x_n\in[a,b]$。并且有一组实数 $\lambda_1,\cdots,\lambda_n\geqslant 0$ 满足 $\sum_{i=1}^{n}\lambda_i=1$，那么则有下列不等式关系成立

$$f\left(\sum_{i=1}^{n}\lambda_i x_i\right)\leqslant\sum_{i=1}^{n}\lambda_i f(x_i)$$

如果函数 f 是凹函数，那么不等号方向逆转。



4、数学推导题（***35%**）

如果你完成了上面一道题，相信你对詹森不等式的理解已经非常充分了，下面这道题将强化你运用它的能力。

在上一次作业中，我们要大家使用拉格朗日乘数法来证明几何-算术均值不等式。下面我们就需要你运用詹森不等式来证明几何-算术均值不等式（关于这个不等式的内容，请你参考上一次作业的描述）

注意：这个不等式的证明方法很多，本题的意思是要求你仅仅使用詹森不等式证明之，如果你采用其它方法（例如把上一次作业的答案重复提交），则会被判定为“答非所问”。

end

end

end

end

end

end

end

end

end

In [ ]: