机器学习降维方法实验报告
PCA Isomap LLE

李梓铉

2018 年 12 月 6 日

指导老师: 孟德宇

1 降维学习背景

在现实中得数据通常很难满足训练样本的密集采样，而且属性维度的增加若满足采样条件，会使得赝本的数字达到天文数字。在高维情形下出现的数据样本稀疏，巨大的矩阵导致的计算距离困难等等，就是机器学习方法面临的严重障碍"维数灾难"（curse of dimensionality）解决维数灾难的重要方法之一是降维（dimension reduction），通过数学变换将高维空间转变为一个低维子空间使得样本密度提高，降低距离计算的复杂度。一般来说有效的降维方法要保持前后的样本之间的距离。降维的方法主要分为两类：第一类也是降维最简单的方法就是基于线性变换的线性降维方法，多维缩放（MDS），主成分分析（PCA）算法均属于此类降维方法。线性降维方法是高维到低维空间的线性映射，但对于现实任务中的很多数据，需要使用非线性映射才能找到适当的低维嵌入。第二类方法就是非线性方法，非线性方法中的第一种是核化线性降维，如核主成分分析（KPCA），使用核函数，将高维特征空间中得数据投影到一个超平面上实现降维。第二种是流形学习（manifold learning），这是一种借鉴拓扑流形概念的降维方法。低维流形嵌入到高维空间的分布虽然很复杂，但局部仍然具有欧式空间的性质，因此可以在局部建议降维映射关系，然后在推广至全局。著名的流形学习方法有等度量映射（Isometric mapping）Isomapping，局部性嵌入（local linear embedding）LLE 等。

本文实验了线性降维方法 PCA，核化线性降维 KPCA，以及流形学习 Isomap 与 LLE 算法。实验使用 python 语言，调用 sci-kit 开源的 python 机器学习库，完成随机生成的高维数据的降维。

2 KPCA 方法

To determine the atomic weight of magnesium via its reaction with oxygen and to study the stoichiometry of the reaction (as defined in 2.1):

$$2Mg + O_2 \longrightarrow 2MgO$$

2.1 Definitions

Stoichiometry The relationship between the relative quantities of substances taking part in a reaction or forming a compound, typically a ratio of whole integers.

Atomic mass The mass of an atom of a chemical element expressed in atomic mass units. It is approximately equivalent to the number of protons and neutrons in the atom (the mass number) or to the average number allowing for the relative abundances of different isotopes.

3 Experimental Data

Mass of empty crucible	7.28 g
Mass of crucible and magnesium before heating	8.59 g
Mass of crucible and magnesium oxide after heating	9.46 g
Balance used	#4
Magnesium from sample bottle	#1

4 Sample Calculation

$$
\begin{aligned}
\text{Mass of magnesium metal} \quad &= 8.59\,\text{g} - 7.28\,\text{g} \\
&= 1.31\,\text{g} \\
\text{Mass of magnesium oxide} \quad &= 9.46\,\text{g} - 7.28\,\text{g} \\
&= 2.18\,\text{g} \\
\text{Mass of oxygen} \quad &= 2.18\,\text{g} - 1.31\,\text{g} \\
&= 0.87\,\text{g}
\end{aligned}
$$

Because of this reaction, the required ratio is the atomic weight of magnesium: 16.00 g of oxygen as experimental mass of Mg: experimental mass of oxygen or $\frac{x}{1.31} = \frac{16}{0.87}$ from which, $M_{\mathrm{Mg}} = 16.00 \times \frac{1.31}{0.87} = 24.1 = 24\,\mathrm{g\,mol^{-1}}$ (to two significant figures).

5 Results and Conclusions

The atomic weight of magnesium is concluded to be $24\,\mathrm{g\,mol^{-1}}$, as determined by the stoichiometry of its chemical combination with oxygen. This result is in agreement with the accepted value.

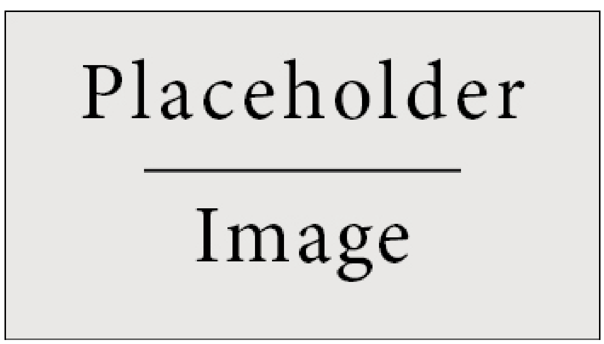

图 1: Figure caption.

6 Discussion of Experimental Uncertainty

The accepted value (periodic table) is $24.3\,\mathrm{g\,mol^{-1}}$ **?**. The percentage discrepancy between the accepted value and the result obtained here is 1.3%.

Because only a single measurement was made, it is not possible to calculate an estimated standard deviation.

The most obvious source of experimental uncertainty is the limited precision of the balance. Other potential sources of experimental uncertainty are: the reaction might not be complete; if not enough time was allowed for total oxidation, less than complete oxidation of the magnesium might have, in part, reacted with nitrogen in the air (incorrect reaction); the magnesium oxide might have absorbed water from the air, and thus weigh "too much." Because the result obtained is close to the accepted value it is possible that some of these experimental uncertainties have fortuitously cancelled one another.

7 Answers to Definitions

a. The *atomic weight of an element* is the relative weight of one of its atoms compared to C-12 with a weight of 12.0000000..., hydrogen with a weight of 1.008, to oxygen with a weight of 16.00. Atomic weight is also the average weight of all the atoms of that element as they occur in nature.

b. The *units of atomic weight* are two-fold, with an identical numerical value. They are g/mole of atoms (or just g/mol) or amu/atom.

c. *Percentage discrepancy* between an accepted (literature) value and an experimental value is

$$\frac{\text{experimental result} - \text{accepted result}}{\text{accepted result}}$$