

Week2

1 UCI数据集

1.1 UCI数据集官网介绍

UCI (University of California Irvine) 数据集是美国加州大学欧文分校提出的一种适合模式识别和机器学习方向的开源数据集，很多学者选择使用UCI上的数据集来验证自己所提算法的正确性。博文写作时已拥有488个数据集，数据集还在不断扩充中，这些数据集主要分为二值分类问题、多分类问题以及回归拟合问题。UCI数据集提供了各个数据集的上主要属性，可以根据自己提出的各类算法在其数据集上做实验结果论证，证明自己所提算法的合理性。

UCI数据集官网地址：<https://archive.ics.uci.edu/ml/index.php>

UCI数据集数据地址：<https://archive.ics.uci.edu/ml/datasets.php>

UCI



Machine Learning Repository

Center for Machine Learning and Intelligent Systems

AboutCitation PolicyDonate a Data SetContact

Search

Repository

Web

View ALL Data Sets

数据集页面入口

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 488 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. For a general overview of the Repository, please visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#).

最新消息

最新数据集

经典数据集

Latest News:

09-24-2018:

Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!

04-04-2013:

Welcome to the new Repository admins Kevin Bache and Moshe Lichman!

03-01-2010:

Note from donor regarding Netflix data

10-16-2009:

Two new data sets have been added.

09-14-2009:

Several data sets have been added.

03-24-2008:

New data sets have been added!

06-25-2007:

Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Featured Data Set: [M. Tuberculosis Genes](#)




Data Type: Relational

Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided

Newest Data Sets:

10-06-2019:

 [WISDM Smartphone and Smartwatch Activity and Biometrics Dataset](#)


09-30-2019:

 [Hepatitis C Virus \(HCV\) for Egyptian patients](#)

09-23-2019:

 [QSAR fish toxicity](#)

09-23-2019:

 [QSAR aquatic toxicity](#)

09-21-2019:

 [Online Retail II](#)

09-20-2019:

 [Human Activity Recognition from Continuous Ambient Sensor Data](#)

09-20-2019:

 [Beijing Multi-Site Air-Quality Data](#)


09-20-2019:

 [MEX](#)

07-30-2019:

 [PPG-DaLiA](#)

07-24-2019:

 [Divorce Predictors data set](#)

07-22-2019:

 [Alcohol QCM Sensor Dataset](#)

07-14-2019:


 [Incident management process enriched event log](#)

Most Popular Data Sets (hits since 2007):

145702:

 [Iris](#)

1735819:

 [Adult](#)

1342508:

 [Wine](#)

1176966:

 [Breast Cancer Wisconsin \(Diagnostic\)](#)

1146070:

 [Heart Disease](#)

141996:

 [Wine Quality](#)

1128866:

 [Car Evaluation](#)

1120010:

 [Bank Marketing](#)

143968:

 [Human Activity Recognition Using Smartphones](#)

1394557:

 [Abalone](#)

1347887:

 [Forest Fires](#)

1392033:

 [Poker Hand](#)

AboutCitation PolicyDonation PolicyContactCML

https://blog.csdn.net/qq_32892383

1

2 核主成分分析 (KPCA)

核主成分分析 (Kernel Principal Component Analysis, KPCA) 是一种非线性数据处理方法，其核心思想是通过一个非线性映射把原始空间的数据投影到高维特征空间，然后在高维特征空间中进行基于主成分分析 (PCA) 的数据处理。

KPCA通常有以下主要应用：降维、特征提取、去噪、故障检测。

2.1 实验

wine数据地址：<http://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data>

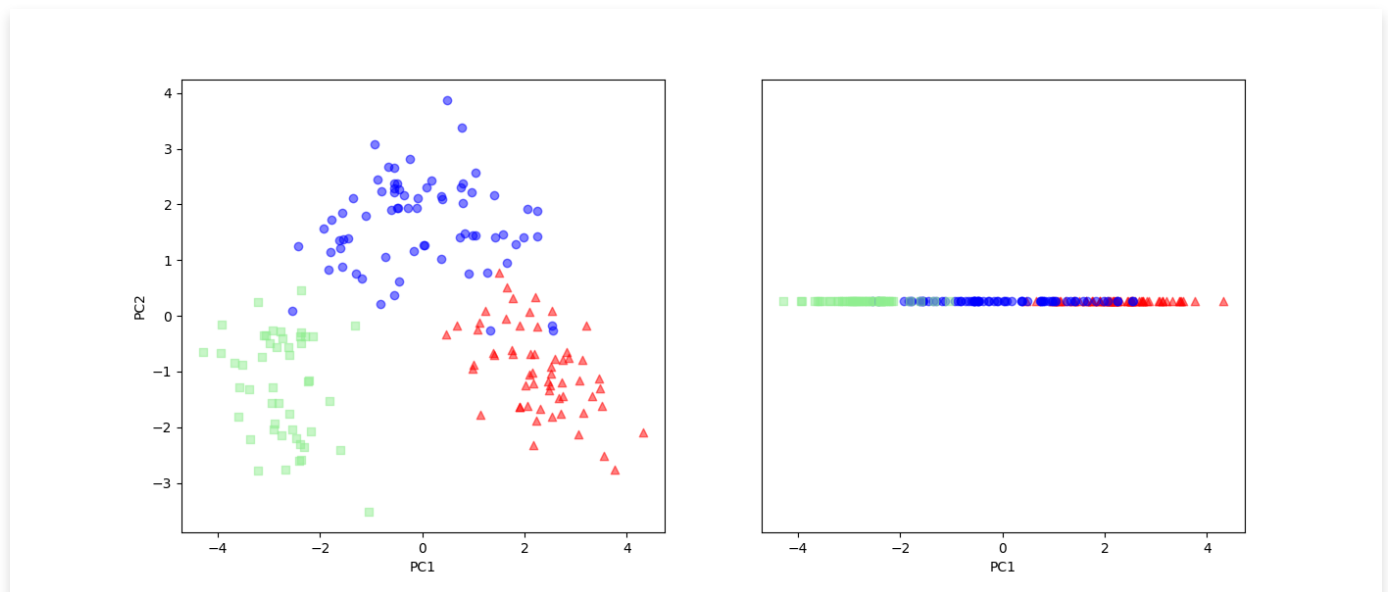


图2.1 wine PCA

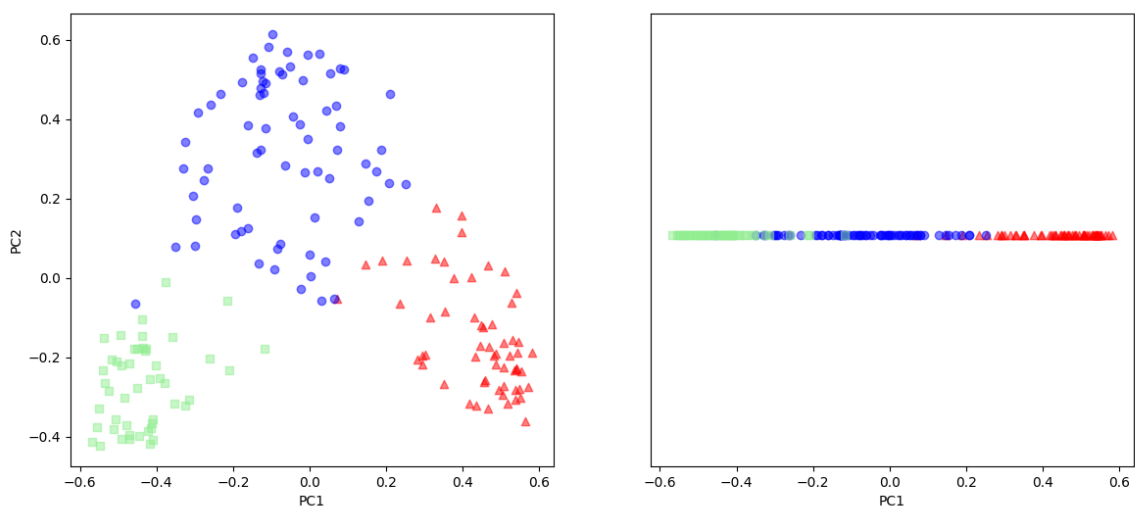


图2.2 wine KPCA

zoo数据地址: <http://archive.ics.uci.edu/ml/machine-learning-databases/zoo/zoo.data>

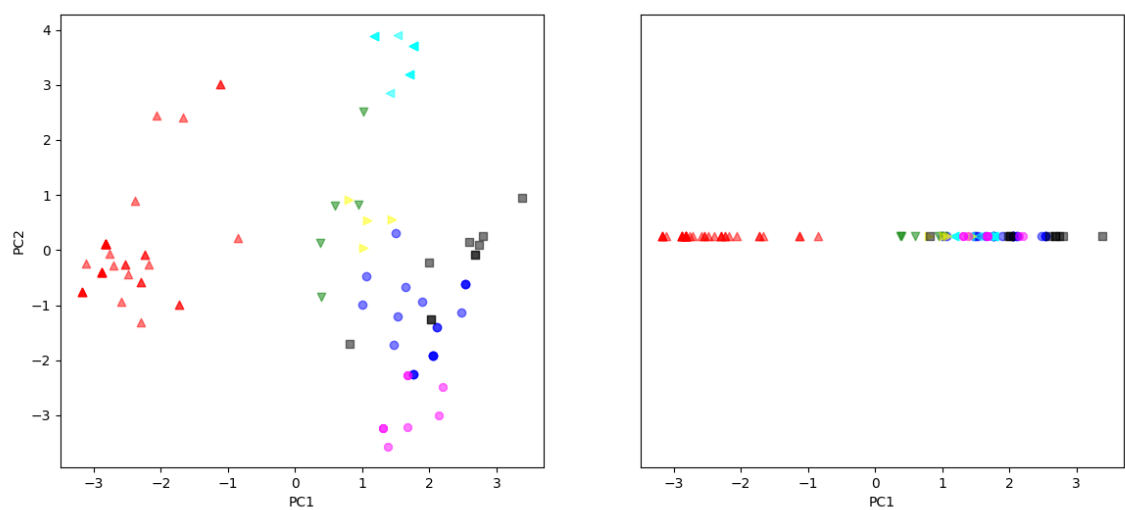


图2.3 zoo PCA

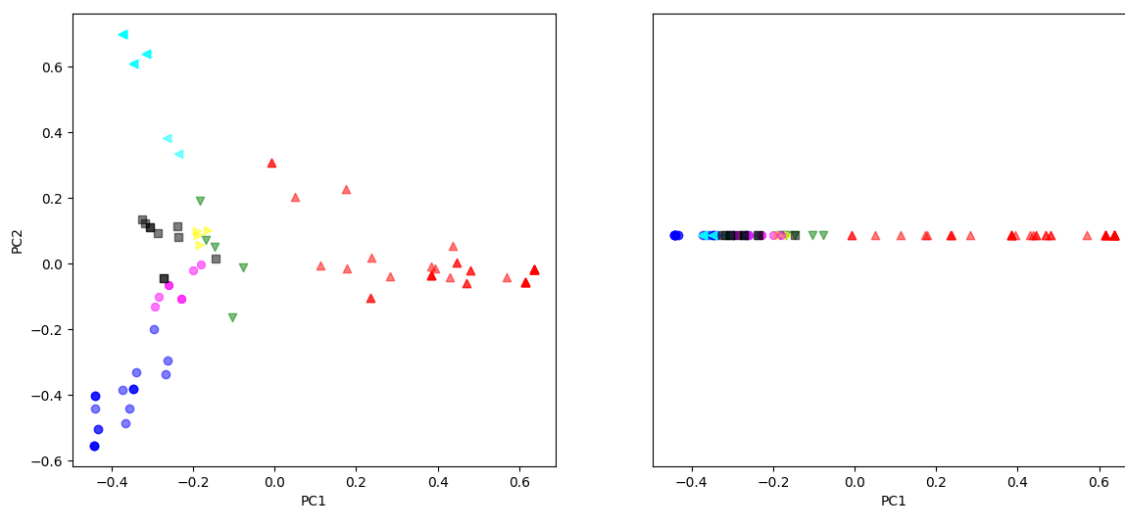


图2.4 zoo KPCA

MNIST数据地址:

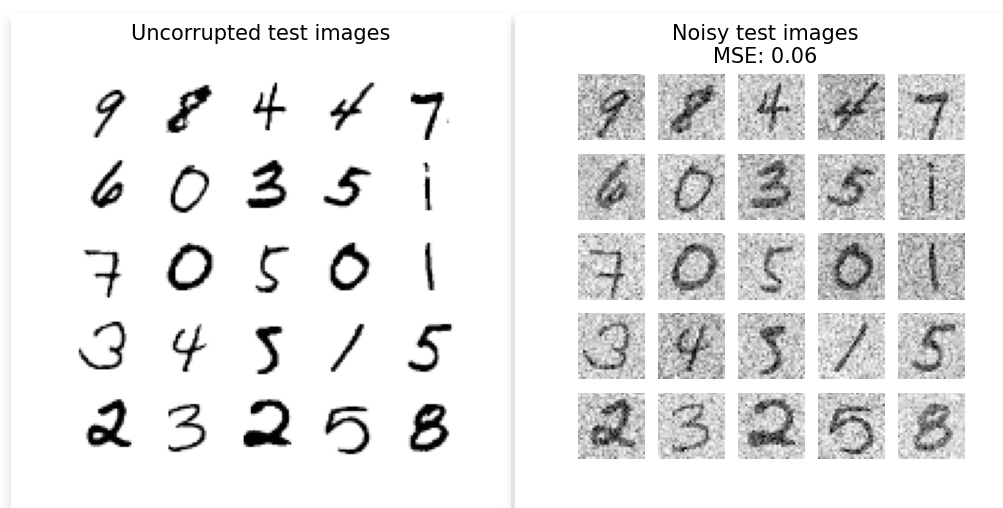


图2.5 noise-free and noisy images

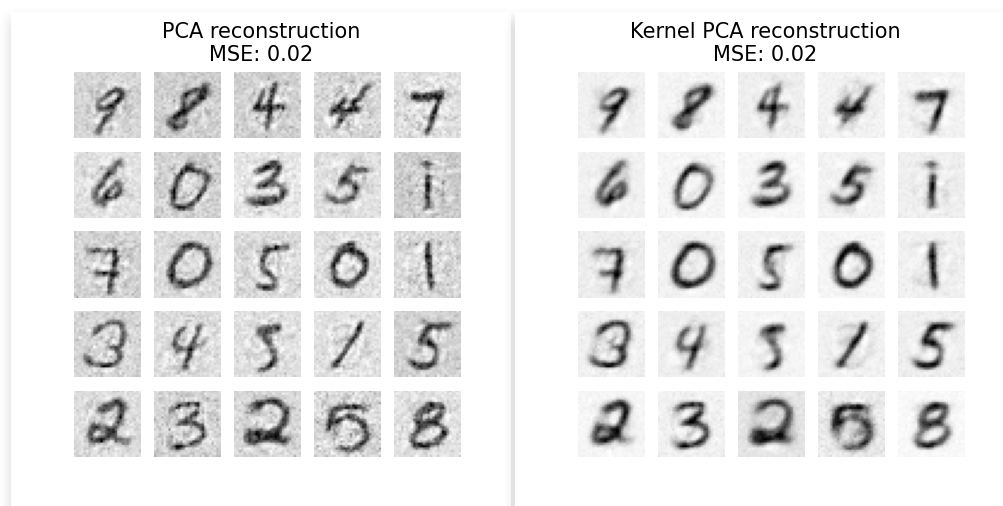


图2.6 denoised imaged with PCA and KPCA

3 核岭回归 (KRR)

WineQuality 数据地址: http://archive.ics.uci.edu/ml/machine-learning-databases/00374/energydata_complete.csv

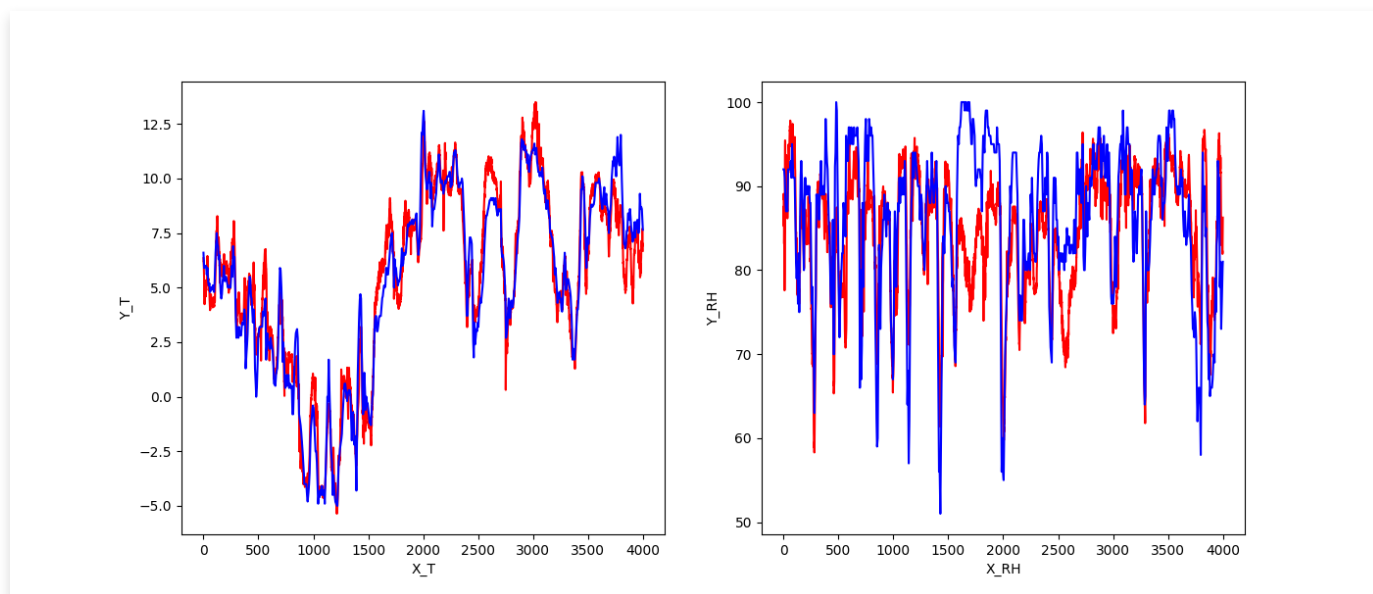


图3.1 WineQuality with KRR

forestfires 数据地址: [http://archive.ics.uci.edu/ml/machine-learning-databases/forest-fires/forest-fires.csv](http://archive.ics.uci.edu/ml/machine-learning-databases/forest-fires/forestfires.csv)

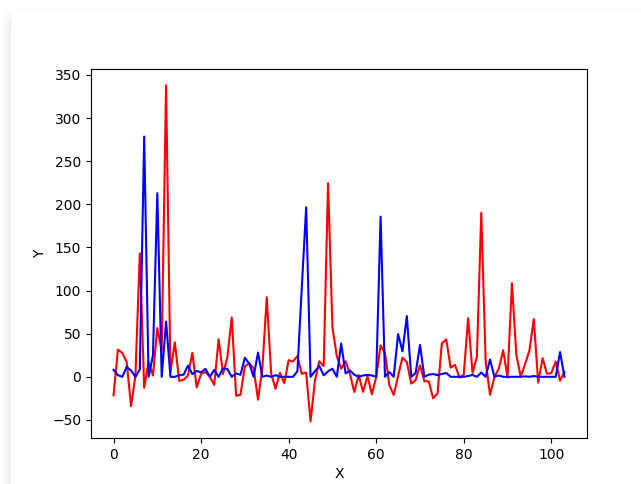


图3.2 forestfires with KRR