第三章 作业

时间	风速（m/s）	功率(MW)
01/01/2016 03:00:00	8.94967	35.971
01/01/2016 03:15:00	8.58567	34.51933
01/01/2016 03:30:00	9.20167m/s	38.435
01/01/2016 03:45:00	9.67667	41.13733
01/01/2016 04:00:00	10.10067	42.351
01/01/2016 04:15:00	9.77767	40.66067
01/01/2016 04:30:00	10.34333	42.91667
01/01/2016 04:45:00	NAN	43.41533
01/01/2016 05:00:00	11.24233	43.847
01/01/2016 05:15:00	11.75767	46.13233
01/01/2016 05:30:00	12.16167	48.19467
01/01/2016 05:45:00	11.20233	46.06533
01/01/2016 06:00:00	10.52533	44.60667
01/01/2016 06:15:00	10.99967	45.14967
01/01/2016 06:30:00	10.94933	45.66333
01/01/2016 06:45:00	10.677	45.00834
01/01/2016 07:00:00	0	-1000
01/01/2016 07:15:00	9.99967	42.41767
01/01/2016 07:30:00	9.63633	42.24667
01/01/2016 07:45:00	8.97967	38.286
01/01/2016 08:00:00	8.48467	33.403

表 1 山西某风电场实测风速和实测发电功率数据

1. 表 1 中的数据，如果没有经过数据预处理就进行数据挖掘的话，会有哪些问题？简述数据预处理的意义和步骤。
2. 请对表 1 中的数据进行数据清洗，说明数据清洗步骤。
3. 请使用表 1 中的数据，回答以下问题：

（a）使用最小-最大规范化将风速和功率值变换到[0.0，1.0]区间。

（b）使用 z 分数规范化变换风速 10.677 m/s 和功率值 42.351 MW。

（c）使用小数定标规范化变换功率值 35.971 MW。

（d）指出对于给定数据，你愿意使用哪种方法，陈述你的理由。

4. 不考虑时序特征，试采用分箱法分别对表 1 中的风速数据和功率数据进行平滑去噪。

5. 对一个 5×2 的二维数据矩阵 X 进行主成分分析，累计方差百分比阈值为 0.8，并用散点图对结果进行可视化。

$$X=\begin{bmatrix}-1 & -1 & 0 & 2 & 0\\ -2 & 0 & 0 & 1 & 1\end{bmatrix}^{T}$$

注意：作业要详细说明计算过程，可编程辅助完成。