Regression Analysis for the Forest Fires Data Set

Goal: Predict the burned area of forest fires, in the northeast region of Portugal

The data set is available at <https://archive.ics.uci.edu/ml/datasets/forest+fires>

Note: In Cortez and Morais (2007), the output 'area' was first transformed with a ln(x+1) function.

Relevant Papers:

[Cortez and Morais, 2007] P. Cortez and A. Morais. A Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimarães, Portugal, pp. 512-523, 2007. APPIA, ISBN-13 978-989-95618-0-9.

資料分析概念:

利用迴歸分析的概念來分析這森林火災資料。

1. Cortez and Morais (2007)建議以ln(x+1)轉換，但是否還有更佳的資料轉換模式。
2. 有哪幾個變量(variables)會是important variables。
3. 如何得到較佳的prediction?

作業內容規範:

本作業視為一個小型的project，報告內容為4-5頁，需要包含資料簡介，分析概念說明，分析結果等。