Exercise 1

The Test Error applying Decision Trees for Classification = 0.9710982658959537.

The RMSE applying Decision Trees for Regression = 0.19319530798283413.

The RMSE applying Logistic Regression = 1.222875848580085.

For the classification, the three most relevant features are feature 21, Feature 15, feature 35.

The running time for 2 cores are 110.153395547s.

The running time for 4 cores are 103.759243884s.

We can see more cores can accelerate the speed of the running.

In this exercise, the speed is very slow when I run the large dataset, so I choose the smaller dataset.

Exercise 2

The two solutions:

Gaussian Generalised Linear Model and Poisson Generalised Linear Model.

The distribution of the training data is similar to the distribution of the two models above, so the two models would be suitable solutions.

Comparison:

	Gaussian	Poisson
AIC	144.01862719931898	249.8425922971907
RMSE	0.2952093388848161	0.549246673699987
training time(s)	23.847788956	14.474407464

The Gaussian Generalised Linear Model has smaller AIC and RMSE than Poisson Generalised Linear Model.

Posisson model has less training time.

Because smaller values of AIC and RMSE indicate better performance, so the Gaussian Generalised Linear Model is the better choice between the two solutions.

In this exercise, I cannot run the whole dataset because it is out of memory, so I choose the smaller dataset.

```
|Exception in thread "dispatcher-event-loop-9"
|java.lang.OutOfMemoryError: Java heap space
```