

Trabalho Predição Mineração de Dados

Pedro Batista - pedro@ufpa.br

October 12, 2010

1 A base de dados

Inicialmente testes realizados na base de dados mostraram resultados muito ruins. Para tentar melhorar o resultado, analisamos o fluxo plotado em função da data, mostrado na Figura 1. Esse gráfico nos mostrou um padrão anual, e que varia de acordo com os meses. Desta forma, plotamos na Figura 2, os dados de forma que o fluxo está somado em cada mês. Onde podemos notar, que o fluxo se repete.

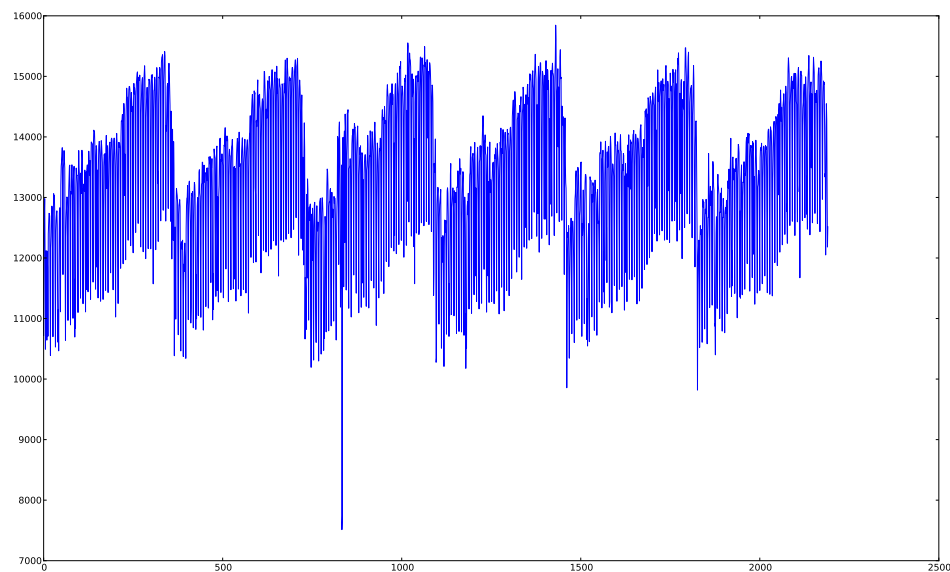


Figure 1: Fluxo plotado em função da data.

2 Resultados

O algoritmo de Rede Neural Multi-camada, foi então aplicado a base de dados de modo que, a informação do dia da semana foi ignorada, assim como a informação do ano. O fluxo original foi somando em cada

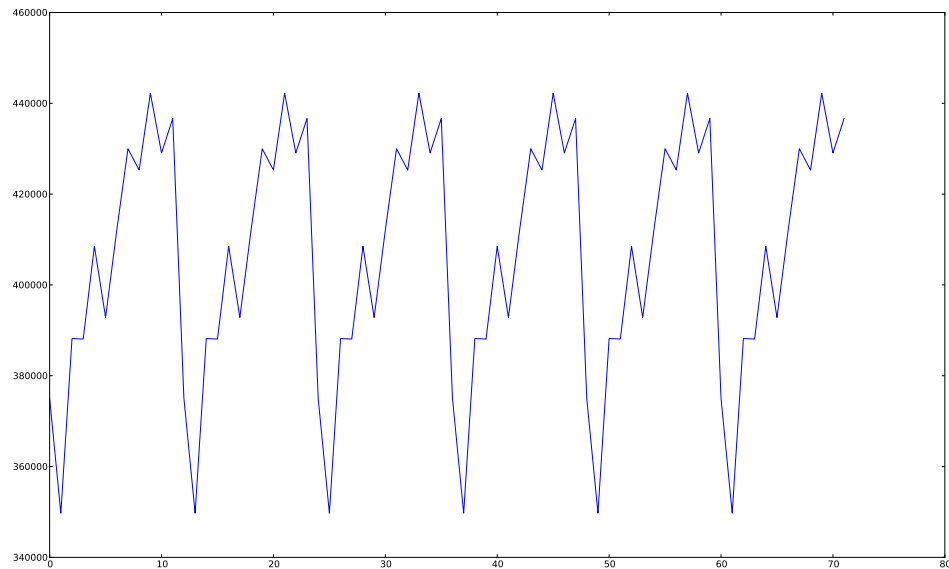


Figure 2: Fluxo somado por mês.

	Mean absolute error	Root squared error	mean Relative error	absolute Root relative squared error
1 mês	4.6171	4.6171	0.0151%	0.0151%
6 meses	0.9032	1.1745	0.0036%	0.0044%
1 ano	2.6945	3.0271	0.0117%	0.0112%

Table 1: Resultados

mês, desta forma, a base conhecida pelo algoritmo continha, apenas a informação número do mês e fluxo correspondente.

Para o teste com 30 dias (testado com 31 dias, mês de dezembro), foi utilizada para teste, os dados de dezembro de 2007 e todo o resto para treino. Para o teste com 6 meses, os últimos seis meses de 2007 foram utilizados para teste e o resto para treino. O teste anual foi realizado com os dados do ano de 2007 e o outros para treino.

Para a RN Multi-camada, foram utilizadas, 20 camadas, taxa de aprendizagem de 0.1 e no máximo 500000 épocas para o treino. Os resultados são mostrados na Tabela 1.