

Relatório Final de Iniciação Científica

Métodos de Otimização aplicados a redes neurais para
detecção de anomalias em transações com cartão de crédito

Bolsista: Frederico José Ribeiro Pelogia

Orientador: Prof. Dr. Luís Felipe Bueno

Número do Projeto FAPESP: 2019/13420-4

Período de vigência: 01/09/2019 a 31/08/2020

Período abordado pelo relatório: 10/02/2020 a 31/08/2020

Departamento de Ciência e Tecnologia
Universidade Federal de São Paulo
Abril de 2020

Resumo do projeto

Neste projeto pretende-se estudar métodos de otimização, em especial algoritmos estocásticos, aplicados ao treinamento de redes neurais. Em um trabalho recente proposto por L. F. Bueno e J. M. Martínez, *On the complexity of solving feasibility problems*, 2018, foi apresentado um algoritmo de primeira ordem com bons resultados de complexidade para problemas de quadrados mínimos. Um dos principais pontos da pesquisa deste projeto será desenvolver uma versão estocástica desse algoritmo. Serão analisados os desempenhos dos algoritmos estudados quando aplicados à detecção de fraudes em operações de cartão de crédito utilizando a base de dados Credit Card Fraud Detection do Kaggle.

Resumo das realizações do período

Este relatório aborda as realizações do período de 10/02/2020 a 31/08/2020. Neste intervalo de tempo, as seguintes atividades foram concluídas:

- Experimentar a aplicação dos algoritmos implementados em Python, sem utilização de bibliotecas prontas de *machine learning*, à base de dados Credit Card Fraud Detection.
- Estudo do método de Levenberg–Marquardt, com base em [TCKleber].
- Estudo teórico de [bmLS].
- Testes numéricos associados a versões estocásticas presentes em [bmLS].
- Aplicações de técnicas de [bmLS] no problema de interesse.

Aplicação dos algoritmos implementados ao problema de detecção de fraudes

- A forma como os métodos estavam implementados não estava adequada para processar os dados do dataset do Kaggle, além de não possibilitar a alteração da estrutura da rede de forma dinâmica.
- A solução foi alterar as implementações para o paradigma de orientação a objeto, construindo uma pequena biblioteca de *Deep Learning* com base na *joelnet*, de Joel Grus.
- NOTA PARA FRED DO FUTURO: Transferir da apresentação de slides para este relatório a explicação sobre os módulos do *joelnet* e as alterações que eu fiz.
- *Falar sobre os bias, que agora estão presentes.
- Após a reformulação das implementações, foi possível testar a capacidade dos algoritmos de detectar as fraudes do banco de dados Credit Card Fraud Detection.
- O método Adam foi o que se destacou para essa tarefa. Obteve o valor de AUCPRC próximo de 0.77 na maioria dos testes .
- *Apresentar comparação dos resultados do Adam do Keras e do implementado sob as mesmas condições de treino e teste.

Estudo do Método de Levenberg–Marquardt

Primeiramente foi feito um estudo sobre alguns métodos clássicos de segunda ordem.

Método de Newton

*falar sobre versão para zero de funções.

A proposta parecida com a do método do gradiente mas com passo proporcional à inversa da matriz hessiana:

$$x_{k+1} = x_k + d_k,$$

onde

$$\nabla^2 f(x) \cdot d_k = -\nabla f(x^k).$$

*talvez colocar teorema de convergência

*explicar depois sobre motivação

Como a matriz Hessiana possui formato $n \times n$, onde n é o número de variáveis, seu cálculo demonstra-se custoso. O método é mais caro ainda, pois além de calcular essa matriz, resolve um sistema com ela. * pesquisar melhor sobre porque Hessianas são caras (argumentos melhores, com *Big O notation*... no slide diz $O(n^3)$ em métodos exatos)

Quasi-Newton: Método da Secante, Barzilai-Borwein e Gauss-Newton

A ideia principal dos métodos Quasi-Newton é herdar a eficácia do Método de Newton utilizando alternativas ou aproximações mais baratas da Hessiana.

Método da Secante

Propõe a escolha de uma matriz M_k tal que

$$M_k s^{k-1} = y^{k-1},$$

onde

$$s^{k-1} = x^k - x^{k-1} \text{ e } y^{k-1} = \nabla f(x^k) - \nabla f(x^{k-1}).$$

Assim, pode-se encontrar d_k tal que

$$M_k d^k = -F(x^k),$$

para que seja possível fazer

$$x^{k+1} = x^k + d^k.$$

Método de Barzilai-Borwein

Sua proposta é ser um intermediário entre o Método do Gradiente, que possui complexidade $O(n)$, e o Método da Secante, que tem complexidade $O(n^2)$. Sendo

$$s^{k-1} = x^k - x^{k-1} \text{ e } y^{k-1} = \nabla f(x^k) - \nabla f(x^{k-1}),$$

A ideia é definir $M_k = \lambda_k I$, implicando que $\lambda_k s^k = y^k$. A melhor aproximação para essa equação é

$$\lambda_k = \frac{(s^k)^T y^k}{\|s^k\|^2}.$$

Então, a direção tomada pelo método será

$$d = -\frac{1}{\lambda_k} \nabla f(x^k),$$

sendo semelhante a uma iteração do Método do Gradiente com passo $\frac{1}{\lambda_k}$. O método também funciona bem em um formato estocástico, se assemelhando, nesse caso, ao SGD com passo $\frac{1}{\lambda_k}$.

Método de Gauss-Newton

Problemas não lineares de quadrados mínimos podem ser formulados da seguinte forma

$$f(x) = \|F(x)\|^2 = \sum_{i=1}^N f_i(x)^2.$$

Neste caso,

$$\nabla^2 f(x) = J_F(x)^T J_F(x) + \sum_{i=1}^N f_i(x) \nabla^2 f_i(x).$$

Uma possível aproximação para a Hessiana é $J_F(x)^T J_F(x)$, tendo em vista que o segundo termo ficará muito pequeno quando próximo do mínimo da função. Assim, a direção de Gauss-Newton é a direção d_k que satisfaz

$$(J_F(x)^T J_F(x)) d_k = -\nabla f(x^k).$$

Segundo [artigoLevenberg], o método é bem definido para quando $J_F(x)^T J_F(x)$ é invertível e, portanto, definida positiva.

Método de Levenberg-Marquardt

Por fim, o Método de Levenberg-Marquardt ...

$$(J_F(x)^T J_F(x) + \lambda I) d_k = -\nabla f(x^k),$$

onde λ pode ser definido de diversas maneiras. Para a implementação realizada, foi considerado

$$\lambda = \|\nabla f(x^k)\|.$$

Note que pequenos valores de λ deixam o método mais parecido com o Método de Newton, enquanto valores altos de λ aproximam o método do Método do Gradiente.

Utilização de recursos da reserva técnica da bolsa

Durante o período abordado por este relatório, não houve utilização dos recursos financeiros da reserva técnica da bolsa.