

IA353 – Redes Neurais (1s2020)

Exercícios Conceituais 1 – EC 1 - Atividade Individual – Peso 5

Data de entrega da resolução (por e-mail): 21/05/2020

Questão 1) (1,0 pontos)

Em virtude das conclusões com embasamento experimental levantadas pelo paper de 2018 (<https://arxiv.org/pdf/1705.08292.pdf>):

The Marginal Value of Adaptive Gradient Methods in Machine Learning

Ashia C. Wilson[‡], Rebecca Roelofs[‡], Mitchell Stern[‡], Nathan Srebro[†], and Benjamin Recht[‡]
{ashia,roelofs,mitchell}@berkeley.edu, nati@ttic.edu, brecht@berkeley.edu

[‡]University of California, Berkeley

[†]Toyota Technological Institute at Chicago

o(a) aluno(a) é convidado(a) a implementar (Sugestão: se baseie no paper Ruder, S. “An overview of gradient descent optimization algorithms”, arXiv:1609.04747v2, 2017) e aplicar o algoritmo ADAM para resolver um sistema linear de equações (nas versões subdeterminada e overdeterminada), comparando o resultado com aquele produzido por uma proposta básica de gradiente descendente estocástico (SGD, do inglês *Stochastic Gradient Descent*). Note que se trata de um problema convexo. Seguem propostas de código em Matlab / Octave para o SGD. Apresente o seu código do ADAM e os resultados gráficos, devidamente interpretados.

```
% SGD para o caso de sistema linear overdeterminado
clear all;
randn('state',0);
N = 10;
Nit = 500;
X = randn(N,2);
S = sign(randn(N,1));
w = (X'*X)\X'*S;
disp('Optimal solution');
disp(w);
w1 = zeros(2,1);
passo = 0.1;
for it=2:Nit,
    w1(:,it) = w1(:,it-1) - (passo/sqrt(it))*(X'*X*w1(:,it-1)-X'*S);
end
figure(1);
title('Stochastic Gradient Descent');
for it = 1:(Nit-1),
    plot([w1(1,it);w1(1,it+1)], [w1(2,it);w1(2,it+1)]);hold on;
    plot(w1(1,it),w1(2,it), 'o');
    plot(w1(1,it+1),w1(2,it+1), 'o');
end
hold off;
disp('Obtained solution');
disp(w1(:,Nit));
[S X*w X*w1(:,Nit)]
```

```
% SGD para o caso de sistema linear subdeterminado
clear all;
randn('state',0);
N = 10;
Nit = 500;
X = randn(N,2*N);
S = sign(randn(N,1));
w = (X'/(X*X'))*S;
w1 = zeros(2*N,1);
passo = 0.1;
for it=2:Nit,
    w1(:,it) = w1(:,it-1) - (passo/sqrt(it))*(X'*X*w1(:,it-1)-X'*S);
end
figure(1);
title('Stochastic Gradient Descent');
for it = 1:(Nit-1),
    plot([w1(1,it);w1(1,it+1)],[w1(2,it);w1(2,it+1)]);hold on;
    plot(w1(1,it),w1(2,it),'o');
    plot(w1(1,it+1),w1(2,it+1),'o');
end
hold off;
disp(' [Minimum Norm Solution Obtained solution] ');
disp([w w1(:,Nit)]);
[S X*w X*w1(:,Nit)]
```

Nota 1: As conclusões extraídas desta atividade conceitual não devem ser estendidas diretamente ao caso do treinamento de uma rede neural profunda, em virtude da grande diferença existente entre os problemas de otimização envolvidos.

Nota 2: Investigue o que ocorre para variadas condições iniciais, não se restringindo à sugestão de partir da origem.

Questão 2) (1,0 pontos)

Tomando a definição de *generalized Tikhonov regularization* (https://en.wikipedia.org/wiki/Tikhonov_regularization):

$$\|Ax - b\|_P^2 + \|x - x_0\|_Q^2$$

deduza passo-a-passo a solução ótima \mathbf{x}^* apresentada, para quaisquer matrizes simétricas P e Q . Procure justificar o emprego de um coeficiente de regularização fixo e igual a 1 na formulação acima.

Questão 3) (1,0 pontos)

- (a) Prove que matrizes $n \times n$ que podem ser decompostas na forma $M = N^T N$, com $N \in \mathfrak{R}^{m \times n}$, são matrizes simétricas e semi-definidas positivas.
- (b) Supondo $m > n$ e tomando N de posto completo, prove que M tem que ser definida positiva.
- (c) Usando o conceito de matriz ortogonal, prove que N pode não ser única, para um mesmo M .
- (d) Apresente uma proposta para obter N , dada uma matriz M definida positiva.

Questão 4) (0,5 pontos)

Uma pessoa recebeu um modelo linearizado $f_L(x, y)$ para a função $f(x, y) = x\sqrt{y}$, em torno de um determinado ponto (x_0, y_0) . Essa pessoa agora precisa utilizar a versão linear $f_L(x, y)$, mas não é capaz de se lembrar de todos os coeficientes de $f_L(x, y)$, a qual tem a forma:

$$f_L(x, y) = 2x + py - 8,$$

onde o coeficiente p é uma incógnita. Além disso, a pessoa não se lembra do ponto (x_0, y_0) em torno do qual foi feita a linearização. Mostre a esta pessoa como obter os referidos valores de (x_0, y_0) e p , apresentando os passos necessários e os respectivos valores numéricos. Nota: Expansão de uma função $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ em série de Taylor em torno do ponto $\mathbf{x}^* \in \mathfrak{R}^n$:

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{x}^*) (\mathbf{x} - \mathbf{x}^*) + O(3).$$

Questão 5) (1,0 pontos)

Defina os conceitos de indução, dedução e abdução, recorrendo a textos como em:

<https://larspsyll.wordpress.com/2016/01/25/deduction-induction-abduction/>

de autoria de Lars Jörgen Pålsson Syll (Doutor em História da Economia).

Explique por que treinar uma rede neural está associado a um processo de inferência indutiva. Em seguida, apresente uma vantagem e uma desvantagem de cada um desses três tipos de inferência (em geral, não restrito ao contexto do curso) e dê um exemplo de cada um no contexto de um pesquisador treinando uma rede neural artificial.

Segue um exemplo ilustrativo desta última atividade da questão:

RACIOCÍNIO INDUTIVO → O erro de treinamento está alto e eu preciso baixar. Nas outras vezes em que eu me defrontei com erros de treinamento altos, bastou aumentar o número de neurônios na rede neural e o erro de treinamento baixou. Logo, a minha decisão é aumentar o número de neurônios.

Questão 6) (0,5 pontos)

Ciente de que o aprendizado envolve plasticidade sináptica, procure identificar dois trabalhos científicos da literatura (se restringir a trabalhos publicados em periódicos especializados) que descrevem experimentos capazes de apoiar a eficácia e a durabilidade das intervenções de treinamento cognitivo na melhoria de habilidades cognitivas específicas em seres humanos adultos idosos. Descreva a metodologia adotada e os principais resultados alcançados em cada trabalho. Anexe o PDF desses dois trabalhos ao submeter a resolução deste EC1.