IA353A - Neural Networks
EC1

Rafael Claro Ito
(R.A.: 118430)

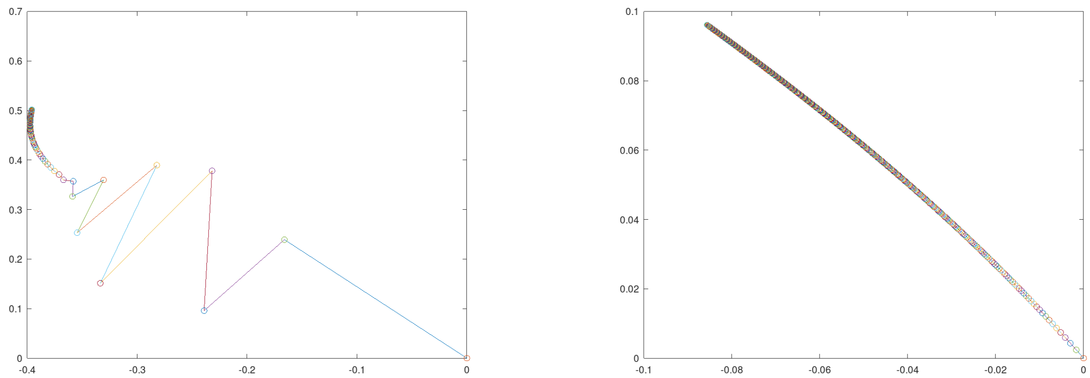
May 2020

Question 1

1.1 SGD

1.1.1 Underdetermined

```
% 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;
passo = 0.001;
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)]
%-----------------------------
% save figure
path = strcat('../figures/Q1/sgd_under_step_', num2str(passo), '.png');
saveas(gcf, path);
```



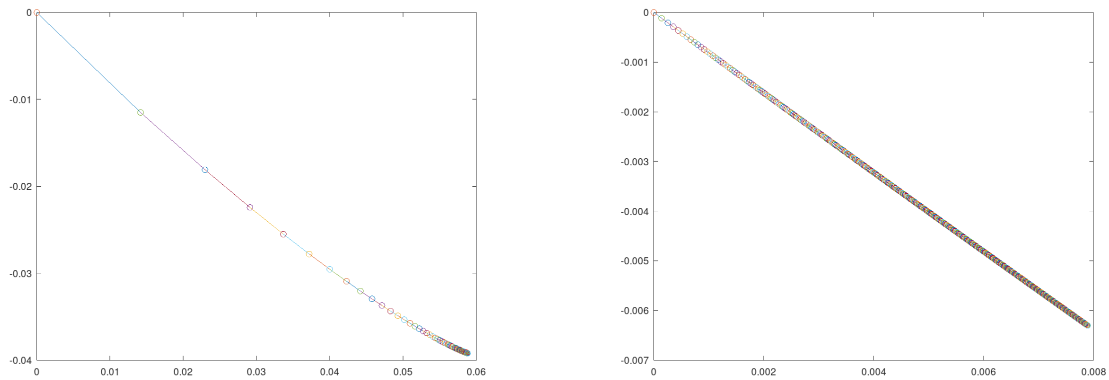
(a) Progression of W using step of 0.1

(b) Progression of W using step of 0.001

Figure 2: Progression of W for underdetermined system using SGD

1.1.2 Overdetermined

```
% SGD para o caso de sistema linear sobredeterminado
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;
passo = 0.001;
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)]
%-----------------------------
% save figure
path = strcat('../figures/Q1/sgd_over_step_', num2str(passo), '.png');
saveas(gcf, path);
```



(a) Progression of W using step of 0.1

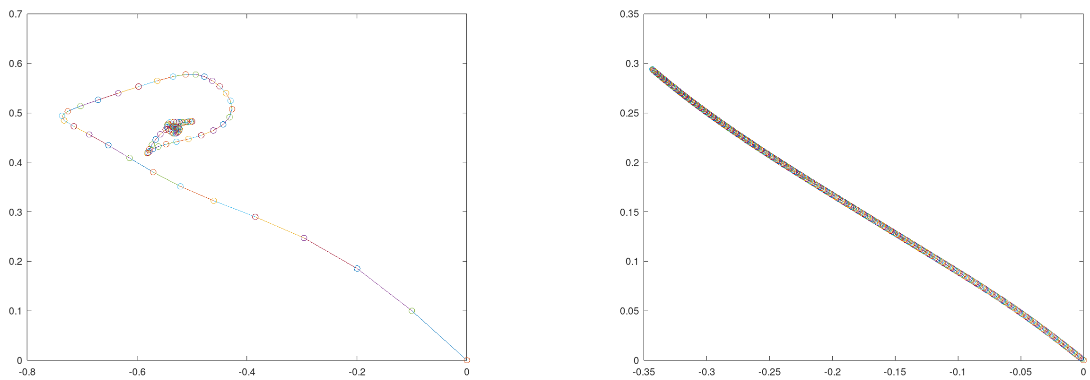
(b) Progression of W using step of 0.001

Figure 3: Progression of W for overdetermined system using SGD

1.2 Adam

1.2.1 Underdetermined

```
% Adam 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;       % optimal solution
w1 = zeros(2*N,1);       % initial weights
%passo = 0.1;
passo = 0.001
%------------------------------
% parameters
beta_1 = 0.9;
beta_2 = 0.999;
e = 1e-8;
m = v = m_hat = v_hat = zeros(2*N,1);
%------------------------------
% loop
for it=2:Nit,
    g = X'*X*w1(:,it-1)-X'*S;
    m = beta_1*m + (1-beta_1)*g;
    v = beta_2*v + (1-beta_2)*g.^2;
    m_hat = m / (1 - beta_1^(it-1));
    v_hat = v / (1 - beta_2^(it-1));
    w1(:,it) = w1(:,it-1) - passo./(sqrt(v_hat) + e) .* m_hat;
    %w1(:,it) = w1(:,it-1) - (passo/sqrt(it))*(X'*X*w1(:,it-1)-X'*S);
end
%------------------------------
% plot weights
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)]
%------------------------------
% save figure
path = strcat('../figures/Q1/adam_under_step_', num2str(passo), '.png');
saveas(gcf, path);
```



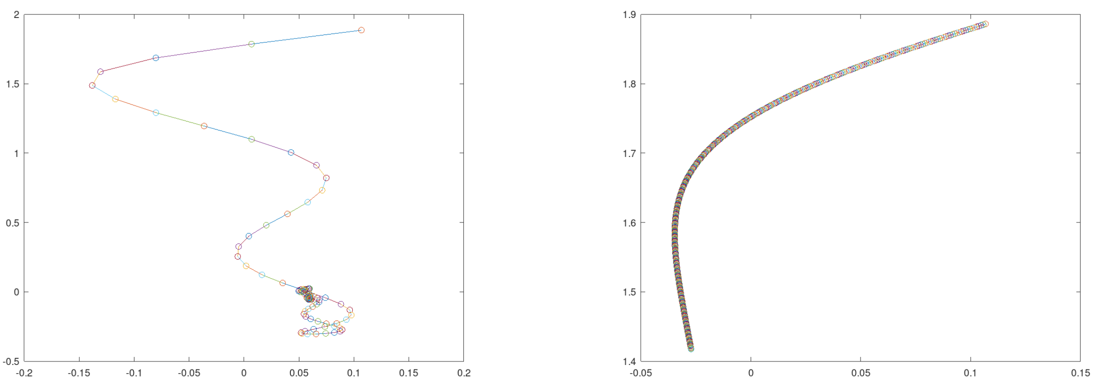
(a) Progression of W using step of 0.1

(b) Progression of W using step of 0.001

Figure 4: Progression of W for underdetermined system using Adam

1.2.2 Overdetermined

```
% Adam para o caso de sistema linear sobredeterminado
clear all;
randn('state',0);
N = 10;
Nit = 500;
X = randn(N,2);
S = sign(randn(N,1));
w = (X'*X)\X'*S;              % optimal solution
disp('Optimal solution');
disp(w);
w1 = randn(2,1);
passo = 0.1;
%passo = 0.001
%----------------------------
% parameters
beta_1 = 0.9;
beta_2 = 0.999;
e = 1e-8;
m = v = m_hat = v_hat = zeros(2,1);
%----------------------------
% loop
for it=2:Nit,
    g = X'*X*w1(:,it-1)-X'*S;
    m = beta_1*m + (1-beta_1).*g;
    v = beta_2*v + (1-beta_2).*g.^2;
    m_hat = m / (1 - beta_1^(it-1));
    v_hat = v / (1 - beta_2^(it-1));
    w1(:,it) = w1(:,it-1) - passo./(sqrt(v_hat) + e) .* m_hat;
    %w1(:,it) = w1(:,it-1) - (passo/sqrt(it))*(X'*X*w1(:,it-1)-X'*S);
end
%----------------------------
% plot weights
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)]
%----------------------------
% save figure
path = strcat('../figures/Q1/adam_over_step_', num2str(passo), '.png');
saveas(gcf, path);
```



(a) Progression of W using step of 0.1

(b) Progression of W using step of 0.001

Figure 5: Progression of W for overdetermined system using Adam

1.3 Comparison

	λ optimum	
	MSE	Accuracy
coarse search	64	1024
fine search	51.5	1091.8

Table 1: Values of regularization coefficient found in coarse and fine searches

Question 2

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

Since we are searching for x that minimizes the previous expression, we will calculate the derivative with relation to x and set it equal to zero:

$$\frac{d}{dx}(\|Ax - b\|_P^2 + \|x - x_0\|_Q^2) = 0$$

Property used: $\|x\|_Q^2 = x^T Q x$

$$\frac{d}{dx}[\overbrace{(Ax-b)^T P(Ax-b)}^{\|Ax-b\|_P^2} + \overbrace{(x-x_0)^T Q(x-x_0)}^{\|x-x_0\|_Q^2}] = 0$$

Property used: $(M+N)^T = M^T + N^T$

$$\frac{d}{dx}\{[(Ax)^T - b^T]P(Ax-b) + (x^T - x_0^T)Q(x-x_0)\} = 0$$

Property used: $(MN)^T = N^T M^T$

$$\frac{d}{dx}\{[x^T A^T - b^T]P(Ax-b) + (x^T - x_0^T)Q(x-x_0)\} = 0$$

$$\frac{d}{dx}[(x^T A^T PAx - x^T A^T Pb - b^T PAx + b^T Pb) + (x^T Qx - x^T Qx_0 - x_0^T Qx + x_0^T Qx_0)] = 0$$

Properties used:

- $\frac{d}{dy}(y^T My) = M^T y + My$.
- $\frac{d}{dy}(y^T My) = 2My$, if $M = M^T$ (i.e. M is symmetric)
- $\frac{d(Ax)}{dx} = A$
- $\frac{d(x^T A)}{dx} = A^T$
- obs.: $A^T PA$ is symmetric, since $(A^T PA)^T = (PA)^T(A^T)^T = A^T P^T A$, but $P^T = P$, since P is symmetric. Then $A^T PA = (A^T PA)^T$

Using the previous properties, we have:

$$\frac{d}{dx}[(x^T A^T PAx - x^T A^T Pb - b^T PAx + b^T Pb) + (x^T Qx - x^T Qx_0 - x_0^T Qx + x_0^T Qx_0)] = 0$$
$$[2A^T PAx - (A^T Pb)^T - b^T PA + 0] + [2Qx - (Qx_0)^T - x_0^T Q + 0] = 0$$
$$2A^T PAx - (Pb)^T (A^T)^T - b^T PA + 2Qx - x_0^T Q^T - x_0^T Q = 0$$
$$2A^T PAx - b^T P^T A - b^T PA + 2Qx - x_0^T Q - x_0^T Q = 0$$
$$2A^T PAx - b^T PA - b^T PA + 2Qx - x_0^T Q - x_0^T Q = 0$$
$$2A^T PAx - 2b^T PA + 2Qx - 2x_0^T Q = 0$$
$$2A^T PAx + 2Qx = 2b^T PA + 2x_0^T Q$$
$$A^T PAx + Qx = b^T PA + x_0^T Q$$
$$(A^T PA + Q)^{-1}(A^T PA + Q)x = (A^T PA + Q)^{-1}(b^T PA + x_0^T Q)$$
$$x = (A^T PA + Q)^{-1}(b^T PA + x_0^T Q)$$
$$x = (A^T PA + Q)^{-1}(b^T P^T A + x_0^T Q^T)$$
$$x = (A^T PA + Q)^{-1}[(Pb)^T A + (Qx_0)^T]$$
$$x = (A^T PA + Q)^{-1}[(A^T Pb)^T + (Qx_0)^T]$$
$$\boxed{x = (A^T PA + Q)^{-1}(A^T Pb + Qx_0)^T}$$

Question 3

a)

b)

c)

We want to prove that N may not be unique in $M = N^T N$. For this, lets consider the orthogonal matrix Q and the new matrix that is the result by its multiplication with N, i.e., QN:

$$M = (QN)^T(QN) = (N^T Q^T)(QN) = N^T(Q^T Q)N$$

Since Q is orthogonal, $Q^T Q$ is equal to the identity matrix I. Hence:

$$M = N^T(Q^T Q)N = N^T I N = N^T N$$

So if M can be decomposed in $N^T N$, it can also be decomposed by $(QN)^T(QN)$, with Q being an orthogonal matrix.

d)

Question 4

The Taylor series expansion of a function $f(x,y)$ in a neighborhood around (x_0,y_0) is as follows:

$$f(x,y) \approx f(x_0,y_0) + \underbrace{f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)}_{\text{first order term}} + ...$$

Ignoring terms from second order and higher, and considering the linearized function given $f_L(x,y) = 2x + py - 8$, we have:

$$f_L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

Taking $f(x,y) = x\sqrt{y}$, we have:

- $f_x(x,y) = \sqrt{y}$
- $f_y(x,y) = \frac{x}{2\sqrt{y}}$

$$2x + py - 8 = \underbrace{x_0\sqrt{y_0}}_{f(x_0,y_0)} + \underbrace{\sqrt{y_0}}_{f_x(x_0,y_0)}(x-x_0) + \underbrace{\frac{x_0}{2\sqrt{y_0}}}_{f_y(x_0,y_0)}(y-y_0)$$

$$2x + py - 8 = x_0\sqrt{y_0} + x\sqrt{y_0} - x_0\sqrt{y_0} + \frac{x_0 y}{2\sqrt{y_0}} - \frac{x_0 y_0}{2\sqrt{y_0}}$$

$$2x + py - 8 = \sqrt{y_0}x + \left(\frac{x_0}{2\sqrt{y_0}}\right) y - \left(\frac{x_0 y_0}{2\sqrt{y_0}}\right)$$

Now, if we compare the terms that only depend on x, that only depend on y and the independent terms (that only depend on x_0 and y_0), we have:

- (i) $2 = \sqrt{y_0} \implies \boxed{y_0 = 4}$
- (ii) $p = \left(\frac{x_0}{2\sqrt{y_0}}\right)$
- (iii) $8 = \left(\frac{x_0 y_0}{2\sqrt{y_0}}\right)$

Substituting in (i) in (iii):

$$\frac{x_0 y_0}{2\sqrt{y_0}} = 8$$

$$\frac{4x_0}{2 \cdot 2} = 8$$

$$\boxed{x_0 = 8}$$

Substituting in (ii):

$$p = \frac{x_0}{2\sqrt{y_0}}$$

$$p = \frac{8}{2 \cdot 2}$$

$$\boxed{p = 2}$$

So the point from where the function was linearized is $(x_0,y_0) = (8,4)$ and the coefficient $p = 2$, then $f_L(x,y) = 2x + 2y - 8$.

Question 5

Question 6

O primeiro paper selecionado é de 2002 com 2173 citações. Esse paper foi publicado no JAMA (The Journal of the American Medical Association) que é um jornal da área médica com 48 publicações por ano pela AMA (American Medical Association).

title: Effects of Cognitive Training Interventions With Older Adults - A Randomized Controlled Trial year: 2002 cited by: 2173 publication: JAMA reference:

Ball K, Berch DB, Helmers KF, et al. Effects of Cognitive Training Interventions With Older Adults: A Randomized Controlled Trial. JAMA. 2002;288(18):2271–2281. doi:10.1001/jama.288.18.2271

O segundo paper selecionado é de 2009 e conta com 315 citações. Esse paper foi publicado no jornal acadêmico Alzheimer's & Dementia, que conta com publicações mensais da associação sem fins lucrativos Journal of the Alzheimer's Association.

title: Immediate and delayed effects of cognitive interventions in healthy elderly: A review of current literature and future directions year: 2009 cited by: 315 publication: Alzheimer's & Dementia (Volume 5, Issue 1, January 2009, Pages 50-60) reference:

Papp K V, Walsh S J, Snyder P J. Immediate and delayed effects of cognitive interventions in healthy elderly: a review of current literature and future directions. Alzheimer's and Dementia 2009; 5(1): 50-60. [PubMed]