The Johns Hopkins University

JHU Engineering for Professionals Program

NEURAL NETWORKS: 625-438/605-447

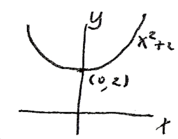
Problem Set #4 Solutions

4.1 Let $y = f(x) = x^2 + 2$

(a) Let the initial guess $x_0 = 3$ and use the method of steepest descent with $\eta=1$ to look for the minimum value of y. Do enough iterations (at least four) to determine the result. Draw a graph.

(b), (c), (d) Repeat part (a) for $\eta = 1/4$, 2, and 0.01, respectively. Discuss the results.

Ans.



$\left\{ \begin{array}{l} x_0 = 3 \\ y = 11 \end{array} \right.$

$f'(x) = 2x$

$x_{n+1} = x_n - \eta f'(x_n) = x_n - 2\eta x_n = (1-2\eta)x_n$

(a)

$\eta = 1$

$f'(3) = 6$

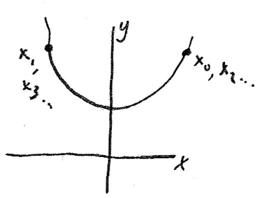
$x_1 = 3 - (1)6 = -3 \quad y = 11$

$x_2 = -3 - (1)(-6) = 3 \quad y = 11$

$\vdots$

$x_3 = -3 \quad y = 11$

$x_4 = 3 \quad y = 11$



Oscillating between $x=+3$ & $x=-3$. Not approaching minimum

(b)

$\eta = \frac{1}{4}$

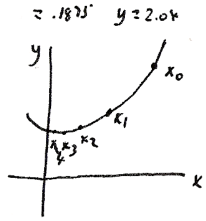
$f'(3) = 6$

$x_1 = 3 - (\frac{1}{4})6 = 1.5 \quad y = 4.25$

$x_2 = 1.5 - (\frac{1}{4})(3) = .75 \quad y = 2.56$

$x_3 = .75 - (\frac{1}{4})(1.5) = .375 \quad y = 2.14$

$x_4 = .375 - (\frac{1}{4})(\frac{3}{4}) = .1875 \quad y = 2.04$



Converging to correct solution, $x=0$

(c)

$\eta = 2$

$f'(3) = 6$

$x_1 = 3 - (2)6 = -9 \quad y = 83$

$x_2 = -9 - (2)(-18) = 27 \quad y = 731$

$x_3 = 27 - (2)(54) = -81 \quad y = 6563$

$x_4 = -81 - (2)(-162) = 243 \quad y = 59051$

Diverging Rapidly

(d)

$\eta = .01$

$f'(3) = 6$

$x_1 = 3 - (.01)6 = 2.94 \quad y = 10.6$

$x_2 = 2.94 - (.01)5.88 = 2.88 \quad y = 10.3$

$x_3 = 2.88 - (.01)5.76 = 2.82 \quad y = 9.95$

$x_4 = 2.82 - (.01)5.64 = 2.76 \quad y = 9.6$

$x_5 = 2.71$

2.66

2.60

2.55

2.50

2.45

$x_{11} = 2.40$

$\vdots$

$x_n = 2.13$

2.085

2.04

$x_{20} = 2.00$

Converging, slowly, to correct solution, $x=0$

Discussion: The solutions to 2.1 illustrate the effect of η, the learning rate. For $\eta = .01$ the process is converging, but too slowly. $\eta = .25$ speeds the process up but increasing further to $\eta = 1$ and $\eta = 2$ drives the solution unstable.

Ans:

Discussion: In (b) + (d) we converge but to a inflection point, not the true minimum. In (a) $\eta = 1$ is really unstable for $x_0 = 4$ and greater. $\eta = 2$ is unstable

(a) $\eta = 1$

$x_0 = 3$

$x_1 = 3 - (1)(3)$

$= 0$

true minimum obtained by accident

But starting at

$x_0 = 4$

$x_1 = 4 - (1)\,16$

$= -12$

$x_2 = -12 - (1)(-2352)$

$= +2340$

unstable

(b) $\eta = 1/4$

$x_0 = 3 \quad y_0 = 2.25$

$x_1 = 3 - (1/4)3$

$= 2.25 \quad y_1 = 1.345$

$x_2 = 2.25 - \left(\frac{1}{4}\right)(.14)$

$= 2.21 \quad y_2 =$

$x_3 = 2.21 - \left(\frac{1}{4}\right)(.097)$

$= 2.19 \quad y_3 = 1.338$

Converging to

$x = 2 \quad y = 1\frac{1}{3}$,

An inflection point, + rate of convergence is decreasing!

(c) $\eta = 2$

$x_0 = 3$

$x_1 = 3 - (2)(3)$

$= -3$

$x_2 = -3 - (2)(-75)$

$= 147$

Going unstable very rapidly

(d) $\eta = .01$

$x_0 = 3$

$x_1 = 3 - (.01)3$

$= 2.97$

2.94

2.91

2.89

will converge very slowly to the inflection point at

$x = 2 \quad y = 1\frac{1}{3}$

as in case (b).

(e) $\eta = .4$

$x_0 = 3 \quad y = 2.25$

$x_1 = 3 - (.4)3$

$= 1.8$

$x_2 = 1.8 - (.4)(.072)$

$= 1.77$

Converging to true minimum at

$x = 0 \quad y = 0$

Note that if $x_k = 1$

$x_{k+1} = 1 - (.4)(1) = .6$

(f) $x_0 = -3 \quad \eta = .25$

$x_0 = -3 \quad y_0 = 74\frac{1}{4}$

$x_1 = -3 - (.25)(-75)$

$= 15\frac{3}{4}$

$x_2 = -728.68$

Diverging

4.2 Repeat problem #4.1 for $y = f(x) = \frac{x^4}{4} - \frac{4}{3}x^3 + 2x^2$

Also, what are the results for

(a) $\eta = 0.4$, $x_0 = 3$ and

(b) $\eta = 0.25$, $x_0 = -3$

Ans:

$$y = f(x) = \frac{x^4}{4} - \frac{4}{3}x^3 + 2x^2 \qquad f'(x) = x^3 - 4x^2 + 4x$$

$$x_{new} = x_{old} - \eta f'(x)$$

	$x_0 = 3$					$x_0 = -3$
	$\eta = 1$	$\eta = 1/4$	$\eta = 2$	$\eta = .01$	$\eta = .4$	$\eta = 1/4$
0	3	3	3	3	3	-3
1	0	2.25	-3			15.75
2		2.21	149			-729
3		2.19	-6,439,333			
4		2.17				
5			Diverge			
6						
7						
8						
9						
10		converging to inflection point at $x = 2$				
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
	$-x^3+4x^2-3x$	$-\frac{1}{4}x^3+x^2$	$-2x^3+8x^2-7x$			$-\frac{1}{4}x^3+x^2$

4.3 Show that for $f(x) = \dfrac{1}{1+e^{-x}}$, $\dfrac{df(x)}{dx} = f'(x) = f(x)\cdot(1 - f(x))$.

Ans:

$$f(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$$

$$f'(x) = \frac{df}{dx} = -(1+e^{-x})^{-2}(e^{-x})(-1)$$

$$= f^2(e^{-x}) \qquad 1+e^{-x} = \frac{1}{f}$$

$$= f^2 \frac{1-f}{f} \qquad \Rightarrow e^{-x} = \frac{1}{f} - 1 = \frac{1-f}{f}$$

$$= f(1-f)$$

4.4 From problem 4.3, for binary data we use the sigmoid function (sometimes called a 'squashing function').

$$f(x) = \frac{1}{1+e^{-x}}$$

and that $f'(x) = f(x)\cdot[1 - f(x)]$.

Now for bi-polar data we must define a *bi-polar* squashing function where

$$g(x) = 2\,f(x) - 1$$

Hence for bi-polar data we must use $g'(x)$ in place of $f'(x)$.
Problem: Show that for $g(x)$ as defined above, we have

$$g'(x) = \frac{1}{2}[1 + g(x)][1 - g(x)]$$

Ans:

$$g(x) = 2f(x) - 1$$

$$g'(x) = 2f'(x) = 2\,f(1-f)$$

$$f = \frac{g+1}{2}$$

So $$g'(x) = 2\left(\frac{g+1}{2}\right)\left(1 - \frac{g+1}{2}\right)$$

$$= (1+g)\left[\frac{2-g-1}{2}\right]$$

$$= \frac{1}{2}(1+g)(1-g)$$

Also

$$g = 2\left(\frac{1}{1+e^{-x}}\right) - 1 = 2\left[1+e^{-x}\right]^{-1} - 1$$

$$g' = 2\left(1+e^{-x}\right)^{-2} e^{-x}$$

$$= 2\left(\frac{g+1}{2}\right)^2 \frac{1-g}{g+1}$$

$$= \frac{1}{2}(1+g)(1-g)$$

$$\frac{g+1}{2} = \frac{1}{1+e^{-x}}$$

$$1+e^{-x} = \frac{2}{g+1}$$

$$e^{-x} = \frac{2}{g+1} - 1$$

$$= \frac{2-g-1}{g+1}$$

$$= \frac{1-g}{g+1}$$

4.5 Use Newton's method to the a zeros (roots) for the following function:

$$f(x) = x^2 - 5x - 10$$

Ans: Note that $f'(x) = 2x - 5$, thus Newton's Method is based on the iterative equation:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^2 - 5x_k - 10}{2x_k - 5}$$

Using a spreadsheet, we get the following iterations:

When we start the process around 2, we converge to -1.53113.

X_{k+1}	Iterative Eq.
2	-14
-14	-6.24242
-6.24242	-2.80059
-2.80059	-1.68314
-1.68314	-1.53389
-1.53389	-1.53113

When we start the process around 5, we converge to 6.53113.

X_{k+1}	Iterative Eq.
5	7
7	6.555556
6.555556	6.531202
6.531202	6.531129
6.531129	6.531129
6.531129	6.531129

Does this make sense? Let's use the quadratic formula to find the roots analytically. Recall, that the roots are $x_r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. Thus, from the original equation $f(x) = ax^2 + bx + c = x^2 - 5x - 10$, we get $\frac{5 \pm \sqrt{25 + 4 \cdot 10}}{2} = \frac{5 \pm \sqrt{65}}{2} = 6.5311288..., -1.5311288...$ in agreement with our final iterates.

4.6 Given $y = f(x) = x^2 + 2$, and using the method of steepest descent, what is the *optimal* value of η, *i.e.,* the value of η that cause the iterations to converge to the minimum value of the function the quickest? Show this analytically.

Ans:

$$y = f(x) = x^2 + 2 \qquad y' = 2x$$

minimum of $y = 2$ at $x = 0$

Our steepest descent equation is

$$x_{i+1} = x_i - \eta f'(x_i)$$

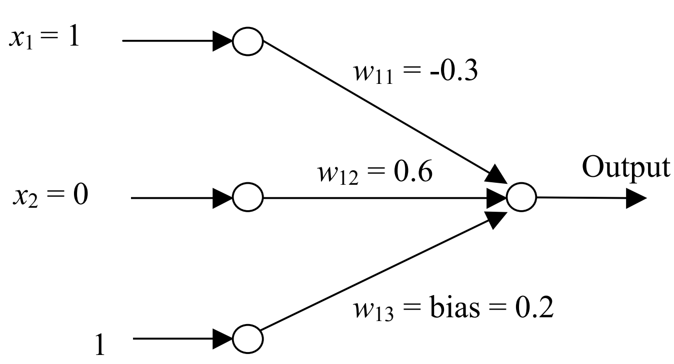
where we want to converge to $x_n \to 0$ as $n \to \infty$

So substitute $0 = x - \eta(2x)$

$1 = 2\eta \Rightarrow \eta = \frac{1}{2}$ will cause convergence in 1 step!

$$\boxed{\text{Optimum } \eta = 0.5 \text{ (independent of } x_0\text{)}}$$

4.7 Using the perceptron delta rule (the steepest descent method for the perceptron) determine the updated weights w_{11} and w_{12} (don't modify the bias) for the given network and values where the activation function is $f(x) = \dfrac{1}{1+e^{-(x+b)}}$, the gain factor value is $\eta = 0.1$, and the desired output is 0.8.



Ans: (Your results may differ somewhat.)

Desired Output	0.8	wij(2)	wij(3)	…	
x1	1	-0.3	-0.2919	-0.28384	-0.27583
x2	0	0.6	0.6	0.6	0.6
f(x)		0.475021	0.477042	0.479053	0.481052
delta_j		0.008104	0.008057	0.00801	0.007962

4.8 In the videos, it was shown that a second-order perceptron can implement the XOR function. Explain how and why it is possible that multiple layers of single-order perceptrons can implement the XOR function.

Short Answer: If one carries out a Taylor Series expansion of the Sigmoid function, it becomes immediately apparent that various terms, in fact, are second order terms. Consider the following:

$$f\left(\sum_k w_{kj} f\left(\sum_j \sum_i w_{ji} x_i\right)\right)$$

where we have a function of a function (the Sigmoid Function). Taking the derivatives of this function f as required in a Taylor Series gives us $f(x)*[1 - f(x)]$ and right away we see that the function is itself squared. Thus any Taylor Series expansion would have second-order terms.

4.9 Write the necessary code to implement the perceptron in 4.2 and using the method of steepest descent, determine the best values for the weights where the initial weights are those given in problem 4.2. Also, indicate what the smallest mean square error is for the best values of the weights after 100 iterations.

Answer: Using an Excel spreadsheet and $\eta = 4$ (for reasonably quick convergence, we obtain after many iterations: Thus, the squared error is $(0.8 - 0.7959)^2 = 0.000016785409$ or thereabouts.

1.141857	1.146591	1.150804	1.154555	1.157898	1.160881
0.6	0.6	0.6	0.6	0.6	0.6
0.792795	0.793572	0.794261	0.794873	0.795418	0.795903
0.004734	0.004212	0.003751	0.003344	0.002983	0.002662