HW 4 this homework need no coding. Put the solution on paper.

All the basic aspects of calculus you need to know for machine learning.

1. Symbolic computation using Chain rule:

Let's define a function called sigmodal activation as

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

(a) Show that

$$\phi'(x) = \phi(x)(1-\phi(x))$$

solution:

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

(b) Now we cascade some functions as

define $y = w_1 x + w_0$

Compute the derivative of $\frac{\partial}{\partial w_1}\phi(w_1, w_0)$ and $\frac{\partial}{\partial w_0}\phi(w_1, w_0)$

define as

$$\phi(y) = \phi(w_1 x + w_0)$$

Express the result in terms of $\phi(y)$ and x using result in (a) to eliminate any derivative of ϕ, i.e, ϕ'.

Solution:

$$\frac{\partial \phi}{\partial w_1} = \phi'(w_1 x + w_0)\frac{\partial}{\partial w_1}(w_1 x + w_0) = \phi'(w_1 x + w_0)x = \phi(1-\phi)x$$

$$\frac{\partial \phi}{\partial w_0} = \phi'(w_1 x + w_0)\frac{\partial}{\partial w_0}(w_1 x + w_0) = \phi'(w_1 x + w_0) = \phi(1-\phi)$$

Note:

Chain rule: we give the chain rule from the elementary calculus.

If there are two function f(x) and g(x), and we composite these two together

$$h = f(g(x))$$

Then the derivative is given by

$$\frac{dh}{dx} = f(g(x)) = f'(g(x))g'(x)$$

2. Gradient calculation:

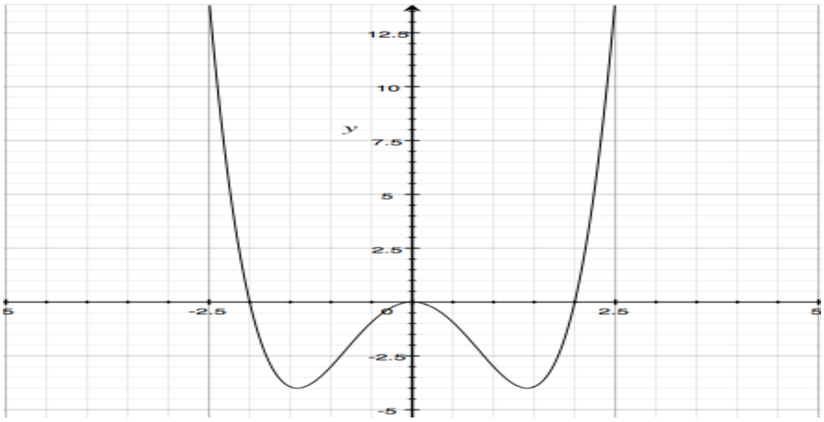
(a) First, you are asked to calculate the derivative of the function

$$g(w) = w^4 - 4w^2$$

Find the minimum of g(w) and plot the function on the interval [-2.5, 2.5].

Note that: in this function, local minimum is the global minimum.

Solution:



$$\frac{dg(w)}{dw} = 4w^3 - 8w = 0$$

the minimum of g(w) = -4 which its w=square root of 2 =√2

(b) Now for a function defined as

$f(w_1, w_2) = (w_1^2 + w_2^2)^2 - 4(w_1^2 + w_2^2)$

compute the gradient $(\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2})$

solution:

$$\frac{\partial f}{\partial w_1} = 4(w_1^2 + w_2^2) \quad * w_1 \quad - 8w_1$$

$$\frac{\partial f}{\partial w_2} = 4(w_1^2 + w_2^2) \quad * w_2 \quad - 8w_2$$

(c) If you run a gradient descent algorithm and start from a random weight coefficient $w_1 \sim r cos\theta, w_2 \sim r sin\theta$, where θ is a random variable with uniform distribution on [0, 2 pi], and r= 2.5 with the gradient in (b) and some reasonable learning rate e.g. η =0.01

$$w_j(n+1) = w_j(n) - \eta \frac{\partial f}{\partial w_j}$$

what would you expect from the trajectory, i.e., $(w_1(n), w_2(n))$ n=1,2,....... will look like from such gradient descent iteration. No coding is necessary. Use some analytic argument to justify your result.

Solution:

The local minimum in this function is a circle defined by r=√2. Therefore, we expect the trajectory will move toward the circle and stop on the circle. The

3. Compute the partial derivative of the function J

(a) $J(w_2, w_1, w_0) = [r - (w_2 x^2 + w_1 x + w_0)]^2$

compute $\frac{\partial J}{\partial w_0}$ and $\frac{\partial J}{\partial w_2}$

solution:

$$\frac{\partial J}{\partial w_0} = -2[r - (w_2 x^2 + w_1 x + w_0)]$$

$$\frac{\partial J}{\partial w_2} = -2[r - (w_2 x^2 + w_1 x + w_0)]x^2$$

(b) $J(C, m, s) = [r - C\exp(-\frac{(x-m)^2}{2s})]^2$

compute$\frac{\partial J}{\partial m}$

solution:

$$\frac{\partial J}{\partial m} = 2[r - Cexp\left(-\frac{(x-m)^2}{2s}\right)](-Cexp\left(-\frac{(x-m)^2}{2s}\right)(-\frac{2(x-m)}{2s})$$

$$=-2[r - Cexp\left(-\frac{(x-m)^2}{2s}\right)]C\,exp\left(-\frac{(x-m)^2}{2s}\right)\frac{(x-m)}{s}$$

(c) $J(w_3, w_2, w_1, w_0) = [r - (w_3 L_3(x) + w_2 L_2(x) + w_1 L_1(x) + w_0)]^2$

where $L_i(x)$ is Legendre polynomial of i^{th} order .

$$L_1(x) = x$$

$$L_2(x) = 1/2(3x^2 - 1)$$

$$L_3(x) = 1/2(5x^3 - 3x)$$

compute $\frac{\partial J}{\partial w_0}$ and $\frac{\partial J}{\partial w_3}$

solution:

$$\frac{\partial J}{\partial w_0} = -2[r - (w_3 L_3(x) + w_2 L_2(x) + w_1 L_1(x) + w_0)]$$

$$\frac{\partial J}{\partial w_1} = -2L_1(x)[r - (w_3 L_3(x) + w_2 L_2(x) + w_1 L_1(x) + w_0)]$$

$$\frac{\partial J}{\partial w_3} = -2L_3(x)[r - (w_3 L_3(x) + w_2 L_2(x) + w_1 L_1(x) + w_0)]$$