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. (30%, each 15%) Symbolic computation using Chain rule:

Let's define a function called sigmodal activation as

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

(a) Show that

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

(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'.

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

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

2. (35%,)Gradient calculation:

(a) (10%) 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.

(b) (10%) 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 $(\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2})$

(c) (15%) If you run a gradient descent algorithm and start from a random weight coefficient $w_1 \sim rcos\theta, w_2 \sim rsin\theta$, 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

$$w_j(n+1) = w_j(n) - \eta \frac{\partial f}{\partial w_1}$$

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.

3. (35%, each 10%) Compute the partial derivative of the function J

(a)(10%) $J(w_2, w_1, w_0) = [r - (w_2x^2 + w_1x + w_0)]^2$

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

(b)(10%) $J(C, m, s) = [r - C\exp(-\frac{(x-m)^2}{2s})]^2$

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

(c)(15%) $J(w_3, w_2, w_1, w_0) = [r - (w_3L_3(x) + w_2L_2(x) + w_1L_1(x) + w_0)]^2$

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 $\frac{\partial J}{\partial w_0}$, $\frac{\partial J}{\partial w_1}$ and $\frac{\partial J}{\partial w_3}$