Assignment 1, Theoretical Part

Charles Huard

February 8, 2019

Question 1

1. The Heaviside function is a piecewise function, with three different value. We therefore need to show that for a particular value of the heaviside function, the derivative of the Relu over the same domain is equal to the Heavyside fonction. Two domains are of interest, $x < 0$ and $x > 0$. Since Relu isn't differientiable at X = 0, we don't need to prove the equality at this value.

   For $x < 0$
   Relu = 0 (and $\frac{\partial 0}{\partial x} = 0$ by definition)
   Heavyside = 0
   Therefore, Heavyside = $\frac{\partial Relu}{\partial x}$ on $x < 0$

   For $x > 0$
   Relu = x, so $\frac{\partial X}{\partial x} = 1$
   Heavyside = 1
   Therefore, Heavyside = $\frac{\partial Relu}{\partial x}$ on $x > 0$

   ■

2. $g(x) = \int H(x)$, by definition since we established that $\frac{\partial g(x)}{\partial x} = H(x)$

   $g(x) = xH(x)$, makes the positive part of H(x) linear, which perfectly mirroirs g(x)

3. H(x) is a 3 piece function, so if we show each piece can be approximate by the sigmoid with a large k, we would show that H(x) can be approximated by a sigmoid with a large k.

   Let N be a large interger.

   For $x < 0$
   $e^{-kx} + 1$ = N, since x is negative

$\frac{1}{e^{-kx}+1} = \frac{1}{N} \approx 0 = H(x)$

For $x = 0$
$\frac{1}{e^{0}+1} = \frac{1}{2} = H(x)$

For $x < 0$
$-kx = -N$, since x is postive
$e^{-N} \approx 0$
$\frac{1}{e^{-kx}+1} \approx 1 = H(x)$

■

4. By the definition that is provided, $F[\phi] = \int\limits_R F(x)\phi(x)dx$
Using integration by parts, we express the derivative to be

$$F'[\phi] = F(x)\phi(x)\Big|_{-\infty}^{\infty} - \int\limits_R F(x)\phi'(x)dx$$

By the definition provided, $\phi(x) = 0$ at ∞ and $-\infty$. We can simply the expression to be

$$F'[\phi] = -\int\limits_R F(x)\phi'(x)dx$$

Which is the desired result.

We then use this definition to express $H'(x) = -\int_{-\infty}^{\infty} H(x)\phi'(x)dx$
By definition H(x) = 0 over $x < 0$. Using this we can reduce the integral to be

$$H'(x) = -\int_0^{\infty} H(x)\phi'(x)dx$$

By definition H(x) = 1 over $x > 0$. Using this we can reduce the integral to be

$$H'(x) = -\int_0^{\infty} \phi'(x)dx = -\Big|_0^{\infty} \phi(x) = -(\phi(\infty) - \phi(0))$$

By definition $\phi(\infty) = 0$

$$H'(x) = -(0 - \phi(0)) = \phi(0)$$

■

Question 2

1. By definition the softmax is

$$\frac{\partial S(x)_i}{\partial x} = \frac{\partial}{\partial x}\frac{exp(x_i)}{\sum_k exp(x_k)}$$

Applying Quotient Rule

$$= \frac{\frac{\partial exp(x)_i}{\partial x_j}(\sum_k exp(x_k)) - \frac{\partial \sum_k exp(x_k)}{\partial x_j} exp(x_i)}{(\sum_k exp(x_k))^2}$$

Refactors to

$$= \frac{\frac{\partial exp(x_i)}{\partial x_j}}{\sum_k exp(x_k)} - \frac{exp(x_i)}{(\sum_k exp(x_k))^2}\frac{\partial}{\partial x_j}(\sum_k exp(x_k))$$

Since $\frac{\partial exp(x_i)}{\partial x_j}$ is 1 if i = j and else, we can express that derivative as being $\delta_{ij} exp(x_i)$. Simlilarly, $\frac{\partial}{\partial x_j}(\sum_k exp(x_k)) = exp(x_j)$
Then

$$= \frac{\delta_{ij} exp(x_i)}{\sum_k exp(x_k)} - \frac{exp(x_i)}{\sum_k exp(x_k)}\frac{exp(x_j)}{\sum_k exp(x_k)}$$

$$= \frac{exp(x_i)}{\sum_k exp(x_k)}(\delta_{ij} - \frac{exp(x_j)}{\sum_k exp(x_k)})$$

By definition of the softmax

$$= S(x_i)(\delta_{ij} - S(x_j))$$

■

2. Knowing $\frac{\partial S(x_i)}{\partial x_j} = S(x_i)(\delta_{ij} - S(x_j))$, distributing we get

$$S(x_i)\delta_{ij} - S(x_i)S(x_j)$$

the left part can be expressed as $diag(S(x_i))$ and the right part can now be expressed as Softmax of a single indice.
We then express the Jacobian matrix as

$$J(S(x)) = diag(S(x)) - S(x)S(x)^T$$

3. First case: i != j

$$\frac{\partial \sigma(x_i)}{\partial x_j} = 0$$

Then if i = j, we need to solve

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

Using the quotient rule

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

Then using simple algebra

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

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

By definition of the sigmoid

$$= \sigma(x)(1-\sigma(x))$$

With both cases we express

$$J(\sigma(x)) = diag(\sigma(x)(1-\sigma(x)))$$

4. We need to show O(n) for the Softmax and the Sigmoid

For the Sigmoid, since $\frac{\partial}{\partial x}\sigma(x)$ is a diagonal matrix, this become a vector multiplication between the diagonal of both matrices, since all other results yields 0. Knowing that the diagonal of an nxn matrix as n elements, we only need to do n multiplication, therefore the multiplication is O(n)

For the Softmax : TODO

Question 3

1.

$$S(x+c) = \frac{e^{x+c}}{\sum_k e^{x+c}}$$

$$= \frac{e^x e^c}{\sum_k e^x e^c}$$

$$= \frac{e^x e^c}{e^c \sum_k e^x}$$

$$= \frac{e^x}{\sum_k e^x}$$

■

2. Proof by contradiction, let S(x) be invariant of scalar multiplication. Then S(x) = S(xc) should hold true.$Let x_1 = 2, x_2 = 4 and c = 2$. Then

$$S(x) = [\frac{e^2}{e^2+e^4}, \frac{e^4}{e^2+e^4}] = [0.1192, 0.88] = S(xc) = [\frac{e^4}{e^4+e^8}, \frac{e^8}{e^4+e^8}] = [0.0179, 0.98]$$

Since the equation is false, we conclude that S(x) is not invariant under scalar multiplication.

We also observe that if $c > 0$, the multiplication by a scalar c raise the value of the most probable class and lowers the other.
If c = 0,$S(x) = \frac{e^0}{\sum_k e^0} = \frac{1}{\sum_j 1}$, which means all class are equally probable and we get an uniform distribution.

3. We must show $\sigma(z) = S(x_2)$ and $1 - \sigma(z) = S(x_1)$ for z being a scalar function of x.

Let $z = x_2 - x_1$

$$\sigma(z) = \frac{1}{1+exp(-z)} = \frac{1}{1+exp(x_1)exp(-x_2)}$$
$$= \frac{exp(x_2)}{exp(x_2)(1+exp(x_1)exp(-x_2)} = \frac{exp(x_2)}{exp(x_2)+exp(x_1)}$$
$$= S(x_2)$$

$$1 - \sigma(z) = 1 - \frac{1}{1+exp(-z)} = 1 - \frac{1}{1+exp(x_1)exp(-x_2)}$$

$$=$$

Question 4

1.