ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

ATHENA Research & Innovation Information Technologies

ΙΕΛ ILSP

M902
Βασικές Μαθηματικές Έννοιες στη Γλωσσική Τεχνολογία

Project Recap

Κυλάφη Χριστίνα-Θεανώ

LT1200012

January, 2021

TABLE OF CONTENTS

Question **1**

1. Let $X = [\ x_1,\ x_2,\ \ldots\ ,\ x_n\ ]$ the feature values of the neural network model and $W = [\ w_1,\ w_2,\ \ldots\ ,\ w_n\ ]$, the learnt weights during the learning process (and w_0 a bias to be added in the weighted sum).

Then the weighted sum is calculated as follows:

$$z = X\ W^T = [\ X\ ]\left[W^T\right] = [\ x_1\ x_2\ \ldots\ x_{n-1}\ x_n\ ]\begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_{n-1} \\ w_n \end{bmatrix}$$

$$= x_1w_1\ +\ x_2w_2\ +\ \ldots\ +\ x_nw_n$$

$$= \sum_1^n x_iw_i$$

where W^T the transpose of the weight matrix.

2. Let σ the sigmoid function, where $\sigma(x) = \dfrac{1}{1+e^{-x}}$.

A bias w_0 is added to the weighted sum z:

$$f(w_1) = \text{bias}\ +\ z = w_0\ +\ \sum_1^n x_iw_i = w_0\ +\ x_1w_1\ +\ x_2w_2\ +\ \ldots\ +\ x_nw_n = x_1w_1\ +\ c\ ,$$

where:
$c = w_0\ +\ x_2w_2\ +\ \ldots\ +\ x_nw_n$

Then:

$$\sigma(f(w_1)) = \frac{1}{1+e^{-f(w_1)}} = \frac{1}{1+e^{-(x_1w_1+c)}} = \frac{1}{1+\frac{1}{e^{(x_1w_1+c)}}} = \frac{1}{\frac{e^{(x_1w_1+c)}+1}{e^{(x_1w_1+c)}}}$$

$$= \frac{e^{(x_1 w_1 + c)}}{e^{(x_1 w_1 + c)} + 1} = \frac{e^{x_1 w_1} e^c}{e^{x_1 w_1} e^c + 1} \qquad (1)$$

$\sigma(f(w_1))$ is differentiable in $Dom(\sigma)$, as a composition of differentiable functions $\sigma,\ f,$ with $\sigma'(f(w_1))$ as follows:

$$\frac{d(\sigma(f(w_1)))}{dw_1} = \sigma'(f(w_1)) = \left(\frac{1}{1 + e^{-f(w_1)}}\right)' \overset{(1)}{=} \left(\frac{e^{x_1 w_1} e^c}{e^{ax} e^b + 1}\right)'$$

$$= \frac{(e^{x_1 w_1} e^c)'(e^{x_1 w_1} e^c + 1)\ -\ (e^{x_1 w_1} e^c)(e^{x_1 w_1} e^c + 1)'}{(e^{x_1 w_1} e^c + 1)^2}$$

$$= \frac{(x_1 e^{x_1 w_1} e^c)(e^{x_1 w_1} e^c + 1)\ -\ (e^{x_1 w_1} e^c)(x_1 e^{x_1 w_1} e^c)}{(e^{x_1 w_1} e^c + 1)^2}$$

$$= \frac{(x_1 e^{x_1 w_1} e^c)(e^{x_1 w_1} e^c + 1\ -\ e^{x_1 w_1} e^c)}{(e^{x_1 w_1} e^c + 1)^2}$$

$$= \frac{x_1 e^{x_1 w_1 + c}}{(e^{x_1 w_1} e^c + 1)^2}$$

Question 2

monotonically increasing -> can be compared

data-> small values ->

underflow

Question **3**

1. The set $\{-4, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ can also be notated as:

$b \rightarrow \{ x \in \mathbb{Z} : -4 \leq x \}$, which means that the set consists of integer numbers that are greater or equal than the integer -4.

2. Let

$A = \{$Greek articles$\} = \{$οι, η, τους, ένας, των, ... $\}$
$B = \{$Greek vowels$\} = \{$ α, ε, η, ι, ο, υ, ω $\}$

Then:
$A \cap B = \{$Greek articles$\} \cap \{$Greek vowels$\}$

$= \{$οι, η, τους, ένας, των, ... $\} \cap \{$ α, ε, η, ι, ο, υ, ω $\}$

$= \{$ α, ε, η, ι, ο, υ, ω $\} = \{$ Greek vowels $\}$

$= B$

3. Let

$A = \{$ Words in Document A $\}$
$B = \{$ Words in Document B $\}$
$A \subset B$

Then:
a. $B - A$ **:** The set consisting of words contained in Document B that are not contained in Document A .

b. $A \subset B \implies |A| \leq |B|$, where $|A|, |B|$ the respective cardinalities of sets A, B .

c. Sets A and B, carry information about the unique words that occur in the respective documents (a set cannot contain duplicates by definition). No information is given about the frequency of each word, thus the total words comprising the documents. Therefore, no safe conclusion can be made concerning the lengths of the two Documents A and B .

Question **4**

Let **2, 4, 15, 8, 7, 7, 6, 3, 3, 3, 40** observations (x_i , $i = 1, 2, \ldots, 11$) of a random variable X.

Then the mean and median values are calculated as follows:

a. Mean value:

$$\frac{\sum_1^{11} x_i}{11} = \frac{2+4+15+8+7+7+6+3+3+3+40}{11} = 8.91$$

Median value:

2, 3, 3, 3, 4, 6 , 7, 7, 8, 15, 40

b. The **mean** value is the most commonly used statistical measure of central tendency with regards to an observation list. However, sometimes due to the structure of the observations' distribution, the **median** is more representative and a more appropriate metric of the data. For example, **not normally distributed** values or the existence of **outliers**, considerably affect the value of the distribution's mean. In our case, there is an **outlier** of value **40** in the dataset, so even though the two values, mean and median, are not too different from one another (**8.91** VS **6**), the **mean** value exhibits **sensitivity** to the **outlier**, **proving** that the **median** is a **better** choice for getting a proper **idea** of the **data distribution** given above.

Question 5

TF - IDF