Considering:

* m = training examples
* n = features
* k = classes
* L = last layer

We have:

*X*=*a*(1), where *a*(1)∈*R*(*m* *x* *n*+1), with bias node.

Thus, *a*(1) rows are the input examples and the columns are the inputs’ features.

*θ*′(1)∈*R*(*n*+1 *x* *k*)

So, *θ*′(1) rows are the weights (parameters) of a specific class and the columns are the classes themselves.

Calculating:

*a*(1)∗*θ*′(1)=*z*(2), where *z*(2)∈*R*(*m* *x* *k*)

Where *z*(2) rows are weighted examples and columns are classes, and each unit of the *z*(2) matrix are like “pre probabilities”, because we haven’t used the activation function yet.

Using activation function:

*a*(2)=*σ*(*z*(2)), where *a*(2)∈*R*(*m* *x* *k*)

Now we have *a*(2) where rows are weighted examples and columns are classes, each unit of the *a*(2) matrix are real probabilities of a *m* example belonging to a *k* class.

Next, we add the bias node to *a*(2), getting *a*(2)∈*R*(*m* *x* *k*+1)

Then, we start repeating the process:

*θ*′(2)∈*R*(*k*+1 *x* *k*)

*a*(2)∗*θ*′(2)=*z*(3), where *z*(3)∈*R*(*m* *x* *k*)

…

When we reach *a*(*L*), which is the last layer, we will have *a*(*L*)∈*R*(*m* *x* *k*), where rows represent examples and the columns represent the classes. Therefore, each unit of this matrix is the probability from 0 to 1 of a example *m* belonging to a class *k*. Consequently, we have to find the maximum probability in each row, then take its index. This index will represent which class our neural network predicted that *m* example to belong. For example, if the highest value of the row is in index 1 (position 2), that *m* example probably belongs to class *k*=2.

Is there any problem with my reasoning?

In lattes:

Considering:

* m = training examples
* n = features
* k = classes
* L = last layer

We have:

$$ X = a^{(1)} $$, where $$ a^{(1)} \in R^{(m \ x\ n+1)} $$, with bias node.

Thus, $$ a^{(1)} $$ rows are the input examples and the columns are the inputs’ features.

$$ \theta'^{(1)} \in R^{(n+1 \ x\ k)} $$

So, $$ \theta'^{(1)} $$ rows are the weights (parameters) of a specific class and the columns are the classes themselves.

Calculating:

$$ a^{(1)} \* \theta'^{(1)} = z^{(2)} $$, where $$ z^{(2)} \in R^{(m \ x\ k)} $$

Where $$ z^{(2)} $$ rows are weighted examples and columns are classes, and each unit of the $$ z^{(2)} $$ matrix are like “pre probabilities”, because we haven’t used the activation function yet.

Using activation function:

$$ a^{(2)} = \sigma (z^{(2)}) $$, where $$ a^{(2)} \in R^{(m \ x\ k)} $$

Now we have $$ a^{(2)} $$ where rows are weighted examples and columns are classes, each unit of the $$ a^{(2)} $$ matrix are real probabilities of a $$ m $$ example belonging to a $$ k $$ class.

Next, we add the bias node to $$ a^{(2)} $$, getting $$ a^{(2)} \in R^{(m \ x\ k+1)} $$

Then, we start repeating the process:

$$ \theta'^{(2)} \in R^{(k+1 \ x\ k)} $$

$$ a^{(2)} \* \theta'^{(2)} = z^{(3)} $$, where $$ z^{(3)} \in R^{(m \ x\ k)} $$

…

When we reach $$ a^{(L)} $$, which is the last layer, we will have $$ a^{(L)} \in R^{(m \ x\ k)} $$, where rows represent examples and the columns represent the classes. Therefore, each unit of this matrix is the probability from 0 to 1 of a example $$ m $$ belonging to a class $$ k $$. Consequently, we have to find the maximum probability in each row, then take its index. This index will represent which class our neural network predicted that $$ m $$ example to belong. For example, if the highest value of the row is in index 1 (position 2), that $$ m $$ example probably belongs to class $$ k = 2 $$.

Is there any problem with my reasoning?