**Learning hypotheses**

**ANN**

In the case of ANN, there are several layers connected, and each layer is composed of several units. The idea is that each layer is connected

* Cum arata weighturile

**Details about the ML models/architectures**

We use SparseCategoricalCrossentropy loss from Keras.. (<https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy>).

This is suited for multi-class classification, and it receives the ground truth labels. We also directly pass the logits for numerical stability.

Mathematically speaking, the following formula is implemented behind the scenes.



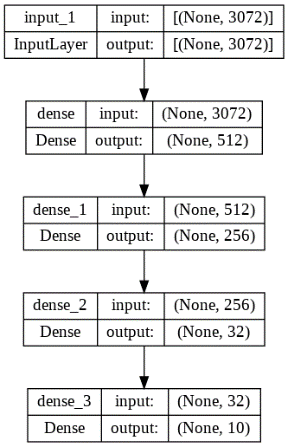
The sparse labels are turned into one-hot encodings, therefore pgtc will be 1 if c is the true class, 0 otherwise and ppredc represents the predicted probability for class c which is computed from the logits.

In our case c takes values from 0 to 9, because there are 10 classes.

We use 2 types of architectures, both having around 1.7 million parameters in order to have a fair comparison.

**Artificial Neural Network**

For the first type of model, we use an ANN with 4 dense layers, as depicted in the image bellow. All dense layers have ReLU activation. The model has 1,713,258 trainable parameters.



Calendar

Description automatically generated

**Experimental results**