Review of the best practice or default values for each problem type with regard to the output layer and loss function

Regression Problem

A problem where you predict a real-value quantity.

- Output (Layer) Configuration: One node with a linear activation unit.
- Loss Function: Mean Squared Error (MSE).

Binary Classification Problem

A problem where you classify an example as belonging to one of two classes.

The problem is framed as predicting the likelihood of an example belonging to class one, e.g. the class that you assign the integer value 1, whereas the other class is assigned the value 0.

- Output (Layer) Configuration: One node with a sigmoid activation unit.
- Loss Function: Cross-Entropy, also referred to as Logarithmic loss.

Multi-Class Classification Problem

A problem where you classify an example as belonging to one of more than two classes.

The problem is framed as predicting the likelihood of an example belonging to each class.

- Output (Layer) Configuration: One node for each class using the softmax activation function.
- Loss Function: Cross-Entropy, also referred to as Logarithmic loss.