

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.

Q: Can we use Sigmoid for multi-class classification problems?

A: *You can also use sigmoid for multi-class classification. When you use softmax, basically you get a probability of each class, (join distribution and a multinomial likelihood) whose sum is bound to be one. In case you use sigmoid for multi class classification, it'd be like a marginal distribution and a Bernoulli likelihood.*

Sum of all softmax units are supposed to be 1. In sigmoid, it's not really necessary. Sigmoid just makes output between 0 to 1. The softmax enforces that the sum of the probabilities of all the output classes are equal to one, so in order to increase the probability of a particular class, softmax must correspondingly decrease the probability of at least one of the other classes.