Forward-Feeding Neural Network (FFNN): Takeaways from Learning Handwritten Digits

STUFF TO KNOW BY HEART - EVEN WHEN DRUNK!

*1. Hidden layer nodes of the Logistic kind serve as **OFF / ON detectors of features that are useful** for the classification goal*

*2. In machines with **large capacity**, weights tend to **specialize** in different tasks; in machines with **small capacity**, weights tend to **multi-task***

This note discusses takeaways from using a 3-layer Forward-Feeding Neural Network (FFNN), with the hidden layer consisting of a set of Logistic nodes, to learn to classify handwritten digits.

1. HIDDEN LAYER NODES AS FEATURE DETECTORS

As the machine learns, the weights adapt in such a way that:

- The hidden layer (layer 2) nodes become better and better **OFF / ON detectors of input images' features / patterns that are relevant** to the task of classifying handwritten digits; such features include dashes, circles, hooks, strokes, etc.
- The output layer (layer 3) nodes become **AND / OR logical combinations of the features detected** by the hidden layer; e.g. a circle on top AND a right-hand-side vertical dash make a number 9.

Both of these are achieved thanks to **Logistic nodes' ability to learn complex AND / OR logical combinations**, ultimately.

2. MACHINE CAPACITY & WEIGHT SPECIALIZATION

It is also noteable that the number of hidden layer nodes has an effect on the capacity of the machine: **the more hidden nodes, the larger the capacity of the machine for learning sophisticated patterns**.

When there are more hidden nodes - hence larger machine capacity - individual weights and hidden nodes will have more room to adapt and will tend to specialize in detecting small local patterns. This will improve the machine's "**attention to detail**" and can help **boost classification performance**. However, the same also increases the **risk of overfitting**.

By contrast, when there are relatively few hidden nodes - hence smaller machine capacity - individial weights and hidden nodes will tend to "multi-task" and detect vaguer, less sharp patterns.