All of these topics are active areas of research.

* The network **initialization** is often overlooked as a source of neural network bugs. Initialization over too-large an interval can set initial weights too large, meaning that single neurons have an outsize influence over the network behavior.
* The key difference between a neural network and a regression model is that a neural network is a composition of many nonlinear functions, called **activation functions**. (See: [What is the essential difference between neural network and linear regression](https://stats.stackexchange.com/questions/259950/what-is-the-essential-difference-between-neural-network-and-linear-regression))

Classical neural network results focused on sigmoidal activation functions (logistic or tanhtanh functions). A recent result has found that ReLU (or similar) units tend to work better because the have steeper gradients, so updates can be applied quickly. (See: [Why do we use ReLU in neural networks and how do we use it?](https://stats.stackexchange.com/questions/226923/why-do-we-use-relu-in-neural-networks-and-how-do-we-use-it)) One caution about ReLUs is the "dead neuron" phenomenon, which can stymie learning; leaky relus and similar variants avoid this problem.

https://stats.stackexchange.com/questions/352036/what-should-i-do-when-my-neural-network-doesnt-learn