



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MLP | | SVM | |
|  | ReLU | Tanh | Linear | Gaussian |
| Average Score | 0.931667 | 0.866349 | 0.869206 | 0.937381 |

For the Multi-layer Perceptron classifier, MLP, the ReLu activation was a much faster and a slightly more accurate tool than the Tanh. In my readings, I see that Tanh suffers from the vanishing gradient problem, as does the sigmoid activation. ReLu, however, processes the data much faster while avoiding the vanishing gradient problem. Recent studies have shown the it matches the training error of Tanh in just 1/6 of the time.  
This causes the weights to not update in the network and limits its ability to be as accurate as possible.

As for SVM, the linear proved to be much faster, but less accurate than the gaussian kernel. The squared exponential is nonparametric and as such has infinite complexity to model the data. This allows it to model very complex relationships.