COMP4702/COMP7703 - Machine Learning
Homework 6 - Neural Networks

Marcus Gallagher

Core Questions

1. Write a short (approx. 2-4 sentences) explanation of why is it useful to have bias weights in a single or multi-layer perceptron (MLP)? Hint: use the example of a single neuron with a linear activation function and generalise to a larger network.

2. Figures 1-3 show some results of training an MLP on a 2D test classification problem. The data and all hyperparameter/training settings are identical apart from the random initialization of weights. Compare the results in a few sentences, specifically in terms of the discriminant function, the training and test error and the expected generalisation performance of the trained model.

3. Figure 4 shows another MLP training example. The setup is identical to Figures 1-3 except a `tanh()` activation function has been used instead of a rectified linear (`relu()`) function. Explain why the discriminant function looks different to those in Figures 1-3.

Extension Questions

4. The softmax activation function:

$$y_i = \frac{\exp o_i}{\sum_k \exp o_k}$$

is often used in MLPs with multiple output units. Consider an MLP with 3 output units. If outputs o_2 and o_3 (before activation), are fixed at 0.2, plot the value of softmax as a function of the other output value, o_1 at an appropriate scale.

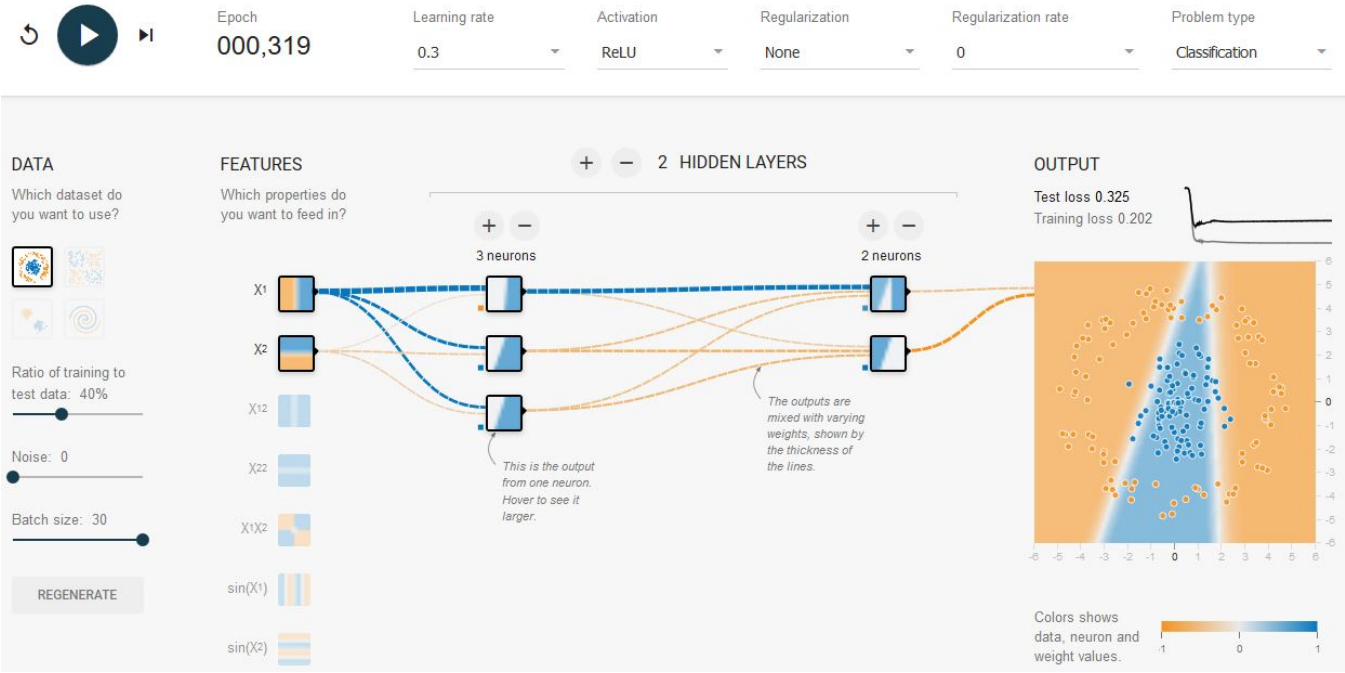

Figure 1: MLP training example.

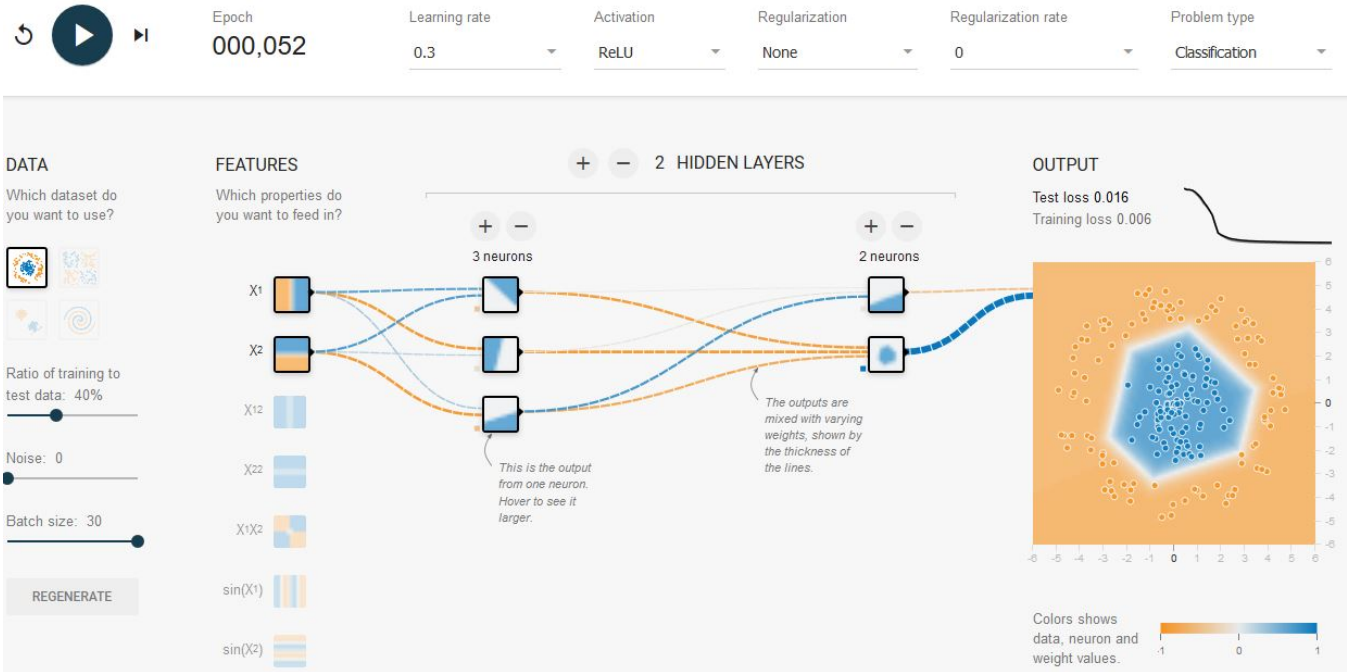

Figure 2: MLP training example.

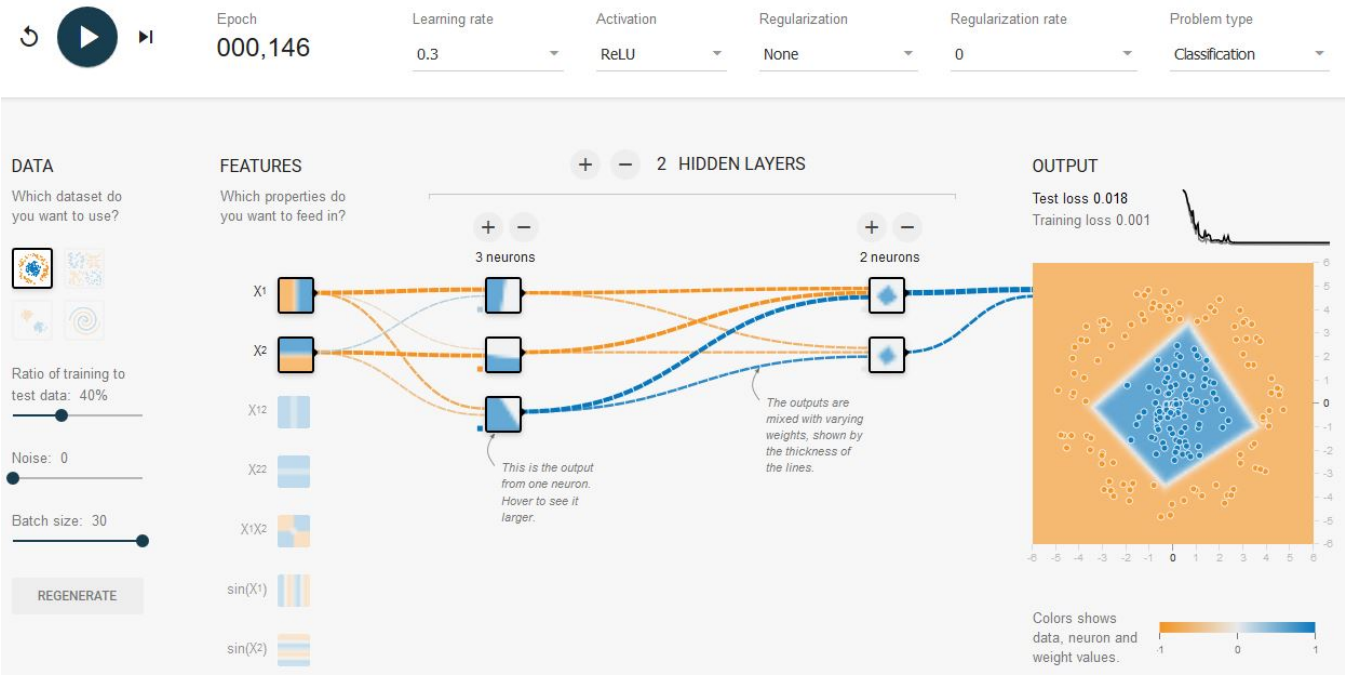

Figure 3: MLP training example.

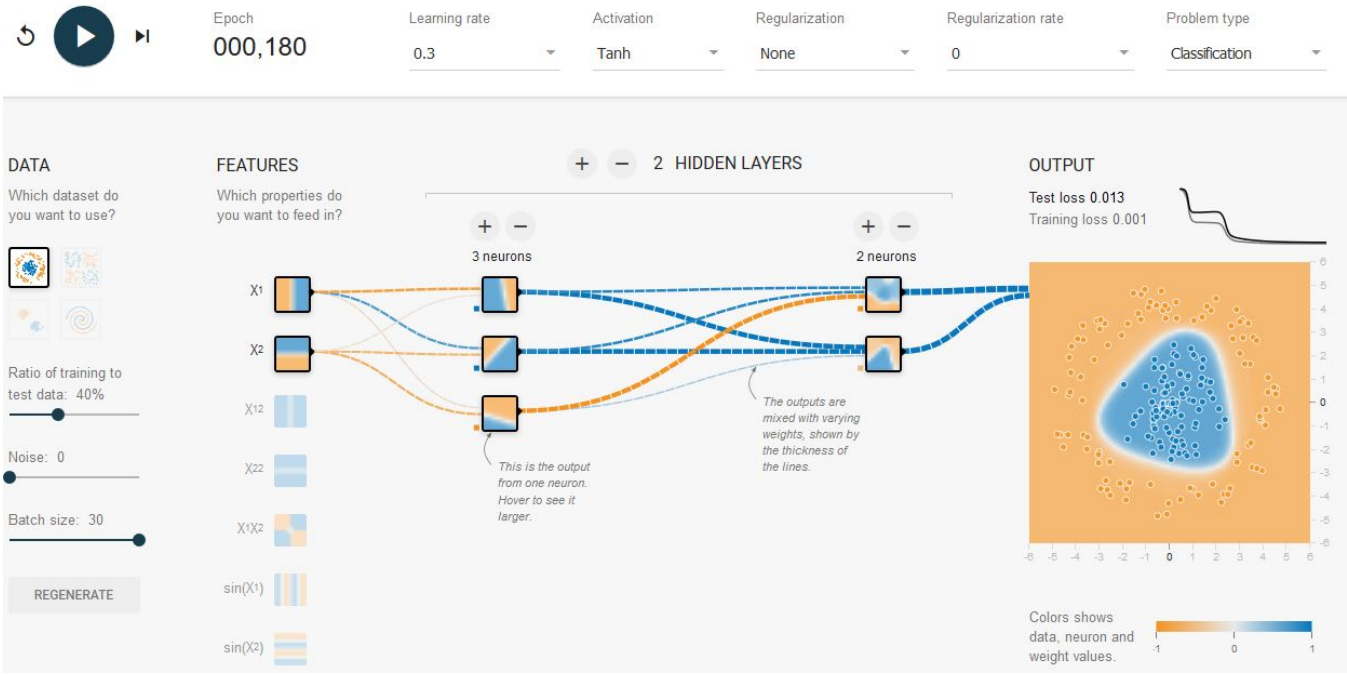

Figure 4: MLP training example.