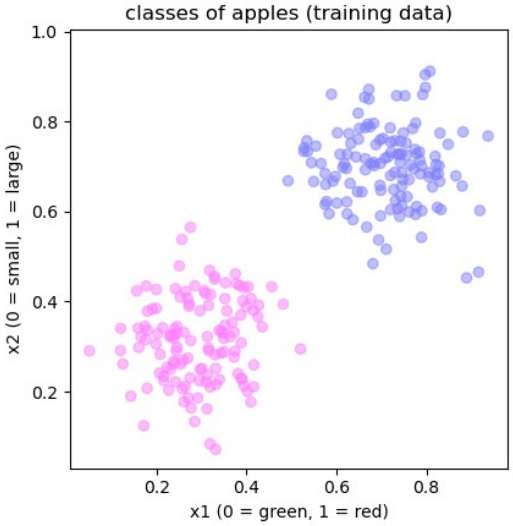

Figure 1. Shows the generated training data for the features x1 and x2 (colour and size)

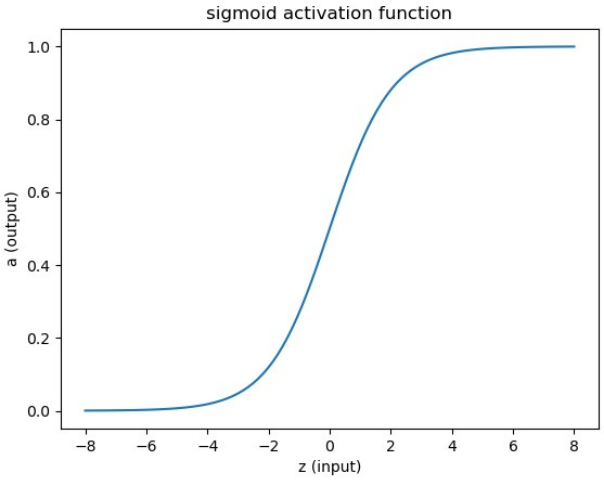

Figure 2. Plot of the sigmoid activation function. sigma(z) = 1 / (1 + e^(-z))

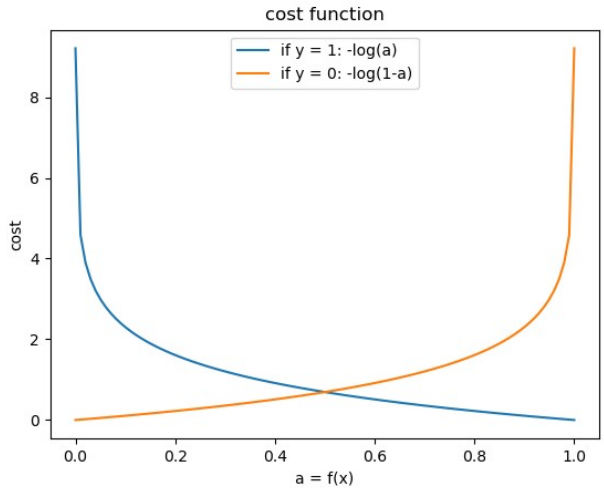

Figure 3. Plot of the cost function. Uses binary cross-entropy loss function.

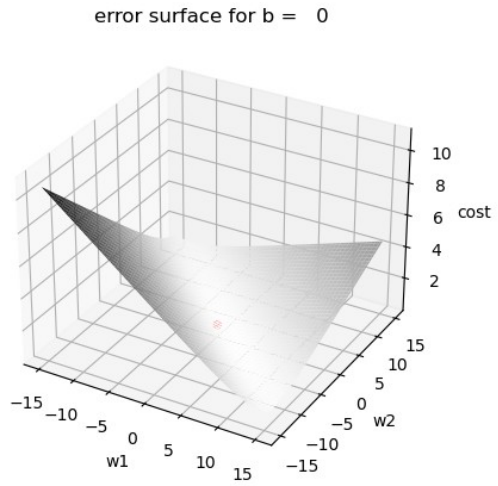

Figure 4. Plot of the error surface across weights 1 and 2. The bias term is user defined.

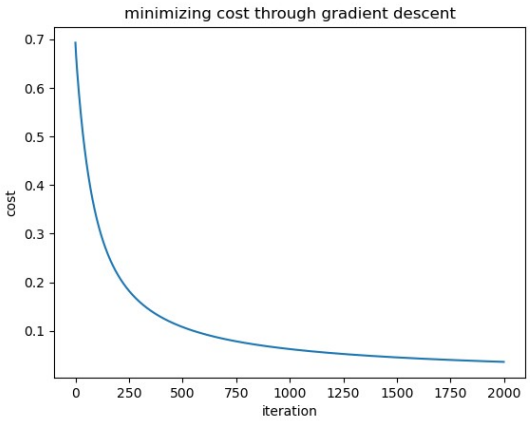

Figure 5. Optimization of the weights using gradient descent. Plot shows a convex cost function converging to a global minimum.

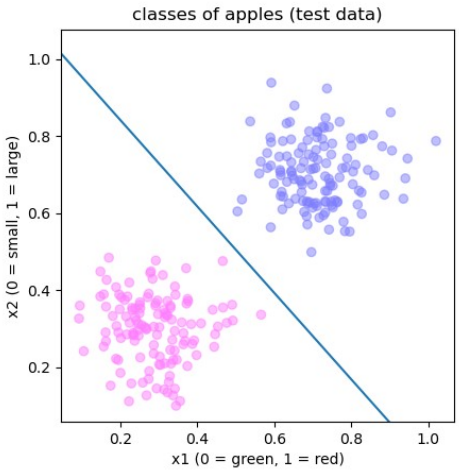

Figure 6. Plot of the test data with the learned decision boundary.