Homework 6: Artificial Neural Networks

Matej Kalc

1. Neural network implementation

I implemented neural networks using equations 1 and 2 for updating the model weights.

$$\frac{\partial J}{\partial W^{(L)}} = A^{(L-1)^T} \cdot (A^{(L)} - Y) \tag{1}$$

$$\frac{\partial J}{\partial W^{(L)}} = A^{(L-1)^T} \cdot (A^{(L)} \otimes (1 - A^{(L)}) \otimes (\frac{\partial J}{\partial A^{(L+1)}} \cdot W^{(L)^T})) \tag{2}$$

Implementation of regression and classification share the majority of the code. The differences are:

- number of outputs,
- in classification we apply softmax in the last layer and
- evaluation of loss (RMSE for regression and log loss for classification).

Additional improvements I implemented are regularization, dropout, gradient descent variations (Polyak and Nesterov GD), and early stopping using a validation set. I tried different activation functions, such as sigmoid, ReLU, tanh, softplus, and perceptron. Sigmoid and tanh were testing the most reliable and gave better results.

2. ANN verification

To verify the correctness of my implementation, I used the equation 3 at the very first iteration. My model passes this test both for regression and classification.

$$|\frac{cost(W_h^L) - cost(W)}{h} - \frac{\partial J}{\partial W_{r,c}^{(L)}}| < \text{tolerance} \tag{3}$$

where W_h is the weight at layer L in row r and column c is increased by h. I check the correctness of the implementation at each first iteration with a tolerance 0.01.

3. Testing Artificial Neural Networks

I tested my implementation on two data sets: housing2r and housing3. The first is a regression problem and the second is a classification problem. For both data sets, I first normalized the data. For the regression problem I compared ANN with Ridge Regression (see Table 1). For both models, I set the regularization term internally.

	Early stopp.	RMSE
ANN with sigmoid activation	Yes	6.65
	No	**6.56**
ANN with tanh activation	Yes	6.91
	No	7.03
Ridge Regression	X	7.67

Table 1. Results of two models on the housing2r data set. The ANN has two layers, the first is of size 13 and the second is of size 6. Regularization term was set internally using 5-Fold cross-validation.

ANN with sigmoid and tanh activation performed better than Ridge Regression. For the second data set I compared ANN with Multinomial Logistic Regression (see Table 2).

	Early stopp.	Accuracy
ANN with sigmoid activation	Yes	**0.91**
	No	0.9
ANN with tanh activation	Yes	0.9
	No	**0.91**
Logistic Regression	X	0.88

Table 2. Results of two models on the housing2r data set. The ANN has one layer with a size 10.

The neural network with both activation functions performed better than the logistic regression in all cases. When using early stop the model stores the best-performing weights when validating. The stored weights are then set as weights of the network. Even taught the model performs more steps with no early stopping, it can have lower accuracy, because the models with early stops use the stored weights.

4. Final prediction

I set the units internally by splitting the train data set and evaluating it onto the split test set. The best units were [50]. The running time was 9:46:38 with an early stop at iteration 84646.