

MSSE 277B: Machine Learning Algorithms

Homework assignment #4: Regression

Assigned Feb. 16 and Due Feb. 28

1. Linear regression using a simple perceptron. (7pt) For getting into graduate school, download the grad school attribute data set from Files.

(a) (2pt) Define the features that are related to predicting the Chance of Admit, and normalize them.

(b) (2pt) Fill in the code for a simple perceptron. Initialize your weights and biases between 0 and 0.05 using your random number generator. Use the mean squares error (MSE) to adjust your weights through back-propagation. Fill in the blanks for the feedforward and backpropagation step.

(c) (3pt) Fill in the code for k-fold validation. Use 80% of the data for training and 20% of the data for testing and do 5-fold validation. Use a learning rate of 0.0001. Are the features good indicators for getting into graduate school? Now remove the GRE scores and do the same test; are GRE scores important?

2. Logistic regression using a simple perceptron. (6pt) For whether your status is sufficient to make it off the Titanic, download the Titanic data set from Files.

(a) (2pt) Process the dataset. What changes from problem 1? Filter out data with missing features. Use one-hot encoders to transform the categorical features and the output survival status.

Hint: Here are some [brief descriptions](#) for the dataset if you are unsure what each descriptor means.

(b) (4pt) Use the simple perceptron model we developed in Q1. Use 80% of the data for training and 20% of the data for testing and do 5-fold validation. Can we predict who will survive? Play around with the features to determine which ones give you a better chance to get back to shore.

3. Nonlinear regression using a simple perceptron and a simple ANN. (7pt) Let's try to fit to a simple sine function $y = 3 \sin(x) + 5$ with the simple perceptron and a simple fully connected network.

(a) (2pt) Use the `generate_data()` function provided in the reference file to generate the training data (5000 points). Do 5-fold cross validation with the simple perceptron model. How well is the prediction? Generate 1000 new points as your test data, and use the `show_correlation()` function to see how well your model agree with the test data.

(b) (3pt) Use the multilayer perceptron regressor of scikit-learn as a simple fully connected neural network. Use one hidden layer with 8 neurons. Do 5-fold cross validation with this simple ANN, and report the MSE on each fold. Visualize the correlation of your model prediction and the true test data. Is the result better than a simple perceptron?

(c) (2pt) Play with the architecture for the ANN, can you improve the performance by including additional hidden layers?