CSE 6363 - *Machine Learning*

Homework 1- Spring 2023

Due Date: March 10 2023

Data Set Generation

This assignment consists a number of implementation and result analysis questions. Some of them (including 1 b), 1 c), 2 require new data. To obtain the data for these problems you need to do the following:

- Go to https://ranger.uta.edu/∼huber/cse6363/Proj1/Proj1gen.php
- Enter your student ID number (the 1000... number on your student ID) and hit submit
- Save the generated web page and submit it with your assignment
- Copy the generated data to files and parameters on your computer and use them with the corresponding questions

Question 3 uses the same data that you used in the first assignment.
Make sure that you enter your own student ID. Results on data for other student ID numbers will not be considered correct solutions.

Linear Regression

1. Consider a simplified fitting problem in the frequency domain where we are looking to find the best fit of data with a set of periodic (trigonometric) basis functions of the form $1,\ x,\ sin^k(k*x)cos(x),\ cos^k(k*x)sin(x),\ sin^{2*k}(2*k*x)cos(x),\ cos^{2*k}(2*k*x)sin(x),\ ...$, where k is effectively the frequency and power increment. The resulting function for a given "frequency increment", k, and "function depth", d (representing the number of sine and cosine terms to be used, and thus determining the number of non-linear features to be used as $2*d+2$), and parameter vector Θ is then:

$$y = \Theta^T\Phi(x) = \Theta_0 * 1 + \Theta_1 * x + \sum_{i=1}^{d}\left(\Theta_{2*i} * \sin^{i*k}(i*k*x)cos(x) + \Theta_{2*i+1} * \cos^{i*k}(i*k*x)sin(x)\right)$$

   For example, if $k=1$ and $d=2$, your basis (feature) functions are $1,\ x,\ sin(x)cos(x),\ cos(x)sin(x),$ $sin^2(2x)cos(x),\ cos^2(2x)sin(x)$, and we are looking for the best matching parameters Θ for the function $\Theta_0+\Theta_1*x+\Theta_2*sin(x)cos(x)+\Theta_3*cos(x)sin(x)+\Theta_4*sin^2(2x)cos(x)+\Theta_5*cos^2(2x)sin(x)$. This means that this problem can be solved using linear regression as the function is linear in terms of the parameters Θ.

   You obtain your value for the "frequency increment" k and thus your basis functions as part of the data generation process described above.

a) Implement a linear regression learner to solve this best fit problem for 1 dimensional data. Make sure your implementation can handle fits for different ”function depths” (at least to ”depth” 6).

b) Apply your regression learner to the data set that was generated for Question 1b) and plot the resulting function for ”function depth” 0, 1, 2, 3, 4, 5, and 6. Plot the resulting function together with the data points (using your favorite plotting program, e.g. Matlab, Octave, ...)

c) Evaluate your regression functions by computing the error on the test data points that were generated for Question 1c) and present the results for the different ”function depths”. Compare the error results and try to determine for what ”function depths” overfitting might be a problem. Which ”function depth” would you consider the best prediction function and why.

Locally Weighted Linear Regression

2. Another way to address nonlinear functions with a lower likelihood of overfitting is the use of locally weighted linear regression where the neighborhood function addresses non-linearity and the feature vector stays simple. In this case we assume that we will use only the raw feature, x, as well as the bias (i.e. a constant feature 1). Thus the locally applied regression function is $y = \Theta_0 + \Theta_1 * x$.

   As discussed in class, locally weighted linear regression solves a linear regression problem for each query point, deriving a local approximation for the shape of the function at that point (as well as for its value). To achieve this, it uses a modified error function that applies a weight to each data point's error that is related to its distance from the query point. Here we will assume that the weight function for the i^{th} data point and query point x is:

$$w^{(i)}(x) = e^{-\frac{(x^{(i)}-x)^2}{2\gamma^2}}$$

   where γ is a measure of the ”locality” of the weight function, indicating how fast the influence of a data point changes with its distance from the query point.

   Your value for γ is provided during data generation.

   a) Implement a locally weighted linear regression learner to solve the best fit problem for 1 dimensional data.

   b) Apply your locally weighted linear regression learner to the data set that was generated for Question 1b) and plot the resulting function together with the data points (using your favorite plotting program, e.g. Matlab, Octave, ...)

   c) Evaluate the locally weighted linear regression on the Test data from Question 1 c). How does the performance compare to the one for the results from Question 1 c) ?

   d) Repeat the experiment and evaluation of part b) and c) using only the first 20 elements of the training data set. How does the performance compare to the one for the results from Question 1 d) ? Why might this be the case ?

   e) Given the results form parts c) and d), do you believe the data set you used was actually derived from a function that is consistent with the function format in Question 1 ? Justify your answer.

Softmax Regression

3. Consider again the problem from Questions 2 in the first assignment where we want to predict the type of material (among 3 material types) of a mug based on four measurements, namely the height, diameter, weight, and hue (color). Assume the same datasets you generated for the first assignment.

   a) Implement Softmax regression to classify this data (use the individual data elements, i.e. height, diameter, weight, and hue, as features). Your implementation should take different data sets as input for learning.

   b) Evaluate the performance of your Softmax regression classifier in the same way as for Homework 1 using leave-one-out validation and compare the results with the ones for KNN (either from your first assignment or, if you did not implement these, using an existing implementation). Present your results and discuss what differences exist and why one method might outperform the others for this problem.

   c) Repeat the evaluation and comparison from part b) when the fourth attribute in the data is removed (i.e. when only the first three features in the input data are available). Again, present the results and discuss what differences exist and why one method might outperform the others in this case.