Machine Learning: Programming Exercise 4

**Neural Networks Learning**

In this exercise, you will implement the backpropagation algorithm for neural networks and apply it to the task of hand-written digit recognition.

**Files needed for this exercise**

* ex4.mlx - MATLAB Live Script that steps you through the exercise
* ex4data1.mat - Training set of hand-written digits
* ex4weights.mat - Neural network parameters for exercise 4
* submit.m - Submission script that sends your solutions to our servers
* displayData.m - Function to help visualize the dataset
* fmincg.m - Function minimization routine (similar to fminunc)
* sigmoid.m - Sigmoid function
* computeNumericalGradient.m - Numerically compute gradients
* checkNNGradients.m - Function to help check your gradients
* debugInitializeWeights.m - Function for initializing weights
* predict.m - Neural network prediction function
* \*sigmoidGradient.m - Compute the gradient of the sigmoid function
* \*randInitializeWeights.m - Randomly initialize weights
* \*nnCostFunction.m - Neural network cost function

***\* indicates files you will need to complete***

**Clear existing variables and confirm that your Current Folder is set correctly**

Click into this section, then click the 'Run Section' button above. This will execute the clear command to clear existing variables and the dir command to list the files in your Current Folder. The output should contain all of the files listed above and the 'lib' folder. If it does not, right-click the 'ex4' folder and select 'Open' before proceding or see the instructions in README.mlx for more details.

clear

dir

**Before you begin**

The workflow for completing and submitting the programming exercises in MATLAB Online differs from the original course instructions. Before beginning this exercise, make sure you have read through the instructions in README.mlx which is included with the programming exercise files. README also contains solutions to the many common issues you may encounter while completing and submitting the exercises in MATLAB Online. Make sure you are following instructions in README and have checked for an existing solution before seeking help on the discussion forums.

**Table of Contents**

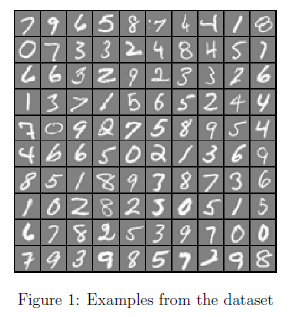
[Neural Networks Learning](#H_18BF401E)  [Files needed for this exercise](#H_4210DE50)  [Clear existing variables and confirm that your Current Folder is set correctly](#H_2D10AFC9)  [Before you begin](#H_9BD8929B)[1. Neural Networks](#H_5D58CA43)  [1.1 Visualizing the data](#H_F8482EE6)  [1.2 Model representation](#H_71B32144)  [1.3 Feedforward and cost function](#H_20D9EF91)  [1.4 Regularized cost function](#H_14FA7ECE)[2. Backpropagation](#H_48811E58)  [2.1 Sigmoid gradient](#H_C0904A30)  [2.2 Random initialization](#H_7BAE046B)  [2.3 Backpropagation](#H_0CCD76C6)  [2.4 Gradient checking](#H_E6862FB4)  [2.5 Regularized neural networks](#H_4E9529DB)  [2.6 Learning parameters using fmincg](#H_0625EF33)[3. Visualizing the hidden layer](#H_CA4CBC73)  [3.1 Optional (ungraded) exercise](#H_D22F158B)[Submission and Grading](#H_A5E37C7C)

**1. Neural Networks**

In the previous exercise, you implemented feedforward propagation for neural networks and used it to predict handwritten digits with the weights we provided. In this exercise, you will implement the backpropagation algorithm to learn the parameters for the neural network.

**1.1 Visualizing the data**

The code below will load the data and display it on a 2-dimensional plot (Figure 1) by calling the function displayData. This is the same dataset that you used in the previous exercise. Run the code below to load the training data into the variables X and y.



load('ex4data1.mat');

m = size(X, 1);

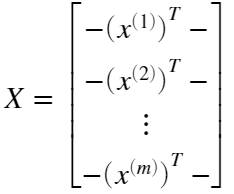
% Randomly select 100 data points to display

sel = randperm(size(X, 1));

sel = sel(1:100);

displayData(X(sel, :));

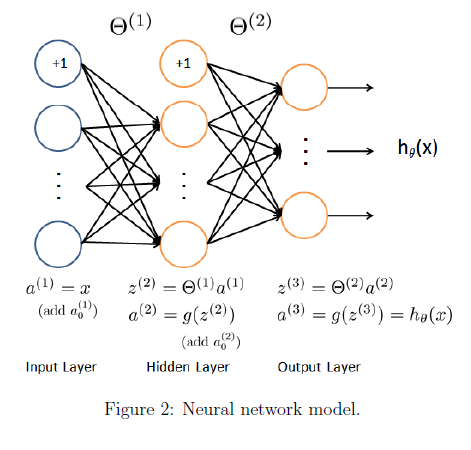
There are 5000 training examples in ex4data1.mat, where each training example is a 20 pixel by 20 pixel grayscale image of the digit. Each pixel is represented by a floating point number indicating the grayscale intensity at that location. The 20 by 20 grid of pixels is 'unrolled' into a 400-dimensional vector. Each of these training examples becomes a single row in our data matrix X. This gives us a 5000 by 400 matrix X where every row is a training example for a handwritten digit image.



The second part of the training set is a 5000-dimensional vector y that contains labels for the training set. To make things more compatible with MATLAB indexing, where there is no zero index, we have mapped the digit zero to the value ten. Therefore, a '0' digit is labeled as '10', while the digits '1' to '9' are labeled as '1' to '9' in their natural order.

**1.2 Model representation**

Our neural network is shown in Figure 2. It has 3 layers- an input layer, a hidden layer and an output layer. Recall that our inputs are pixel values of digit images. Since the images are of size 20 x 20, this gives us 400 input layer units (not counting the extra bias unit which always outputs +1).



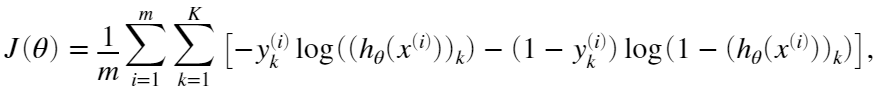
You have been provided with a set of network parameters  already trained by us. These are stored in ex4weights.mat.  Run the code below to load them into Theta1 and Theta2. The parameters have dimensions that are sized for a neural network with 25 units in the second layer and 10 output units (corresponding to the 10 digit classes).

% Load the weights into variables Theta1 and Theta2

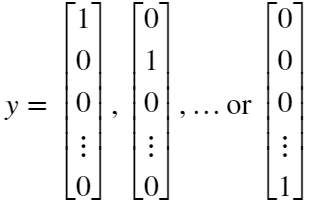
load('ex4weights.mat');

**1.3 Feedforward and cost function**

Now you will implement the cost function and gradient for the neural network. First, complete the code in nnCostFunction.m to return the cost. Recall that the cost function for the neural network (without regularization) is



where  is computed as shown in the Figure 2 and  is the total number of possible labels. Note that  is the activation (output value) of the *k*-th output unit. Also, recall that whereas the original labels (in the variable y) were  for the purpose of training a neural network, we need to recode the labels as vectors containing only values 0 or 1, so that



For example, if  is an image of the digit 5, then the corresponding  (that you should use with the cost function) should be a 10-dimensional vector with , and the other elements equal to 0. You should implement the feedforward computation that computes  for every example *i* and sum the cost over all examples. Your code should also work for a dataset of any size, with any number of labels (you can assume that there are always at least  labels).

**Implementation Note:** The matrix X contains the examples in rows (i.e., X(i,:)' is the i-th training example , expressed as a *n* x 1 vector.) When you complete the code in nnCostFunction.m, you will need to add the column of 1's to the X matrix. The parameters for each unit in the neural network is represented in Theta1 and Theta2 as one row. Specifically, the first row of Theta1 corresponds to the first hidden unit in the second layer. You can use a for loop over the examples to compute the cost. We suggest implementing the feedforward cost *without* regularization first so that it will be easier for you to debug. Later, you will get to implement the regularized cost.

Once you are done, run the code below to call your nnCostFunction using the loaded set of parameters for Theta1 and Theta2. You should see that the cost is about 0.287629.

input\_layer\_size = 400; % 20x20 Input Images of Digits

hidden\_layer\_size = 25; % 25 hidden units

num\_labels = 10; % 10 labels, from 1 to 10 (note that we have mapped "0" to label 10)

% Unroll parameters

nn\_params = [Theta1(:) ; Theta2(:)];

% Weight regularization parameter (we set this to 0 here).

lambda = 0;

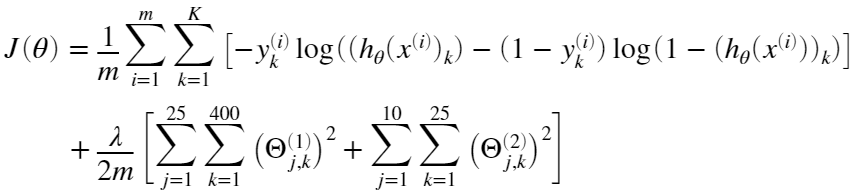
J = nnCostFunction(nn\_params, input\_layer\_size, hidden\_layer\_size, num\_labels, X, y, lambda);

fprintf('Cost at parameters (loaded from ex4weights): %f', J);

*You should now submit your solutions. Enter* **submit***at the command prompt, then enter or confirm your login and token when prompted.*

**1.4 Regularized cost function**

The cost function for neural networks with regularization is given by



You can assume that the neural network will only have 3 layers- an input layer, a hidden layer and an output layer. However, your code should work for any number of input units, hidden units and outputs units. While we have explicitly listed the indices above for  and  for clarity, do note that**your code should in general work with**  and  **of any size.**

Note that you should not be regularizing the terms that correspond to the bias. For the matrices Theta1 and Theta2, this corresponds to the first column of each matrix. You should now add regularization to your cost function. Notice that you can first compute the unregularized cost function *J* using your existing nnCostFunction.m and then later add the cost for the regularization terms. Once you are done, run the code below to call your nnCostFunction using the loaded set of parameters for Theta1 and Theta2, and . You should see that the cost is about 0.383770.

% Weight regularization parameter (we set this to 1 here).

lambda = 1;

J = nnCostFunction(nn\_params, input\_layer\_size, hidden\_layer\_size, num\_labels, X, y, lambda);

fprintf('Cost at parameters (loaded from ex4weights): %f', J);

*You should now submit your solutions. Enter/confirm your login and token in the command window when prompted.*

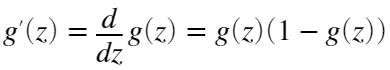
**2. Backpropagation**

In this part of the exercise, you will implement the backpropagation algorithm to compute the gradient for the neural network cost function. You will need to complete the nnCostFunction.m so that it returns an appropriate value for grad. Once you have computed the gradient, you will be able to train the neural network by minimizing the cost function  using an advanced optimizer such as fmincg.

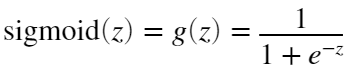
You will first implement the backpropagation algorithm to compute the gradients for the parameters for the (unregularized) neural network. After you have verified that your gradient computation for the unregularized case is correct, you will implement the gradient for the regularized neural network.

**2.1 Sigmoid gradient**

To help you get started with this part of the exercise, you will first implement the sigmoid gradient function. The gradient for the sigmoid function can be computed as



where



When you are done, try testing a few values by calling sigmoidGradient(z) below. For large values (both positive and negative) of z, the gradient should be close to 0. When z = 0, the gradient should be exactly 0.25. Your code should also work with vectors and matrices. For a matrix, your function should perform the sigmoid gradient function on every element.

% Call your sigmoidGradient function

sigmoidGradient(0)

*You should now submit your solutions. Enter* **submit***at the command prompt, then enter or confirm your login and token when prompted.*

**2.2 Random initialization**

When training neural networks, it is important to randomly initialize the parameters for symmetry breaking. One effective strategy for random initialization is to randomly select values for  uniformly in the range . You should use \*. This range of values ensures that the parameters are kept small and makes the learning more efficient.

Your job is to complete randInitializeWeights.m to initialize the weights for Θ; modify the file and fill in the following code:

% Randomly initialize the weights to small values

epsilon\_init = 0.12;

W = rand(L out, 1 + L in) \* 2 \* epsilon\_init - epsilon init;

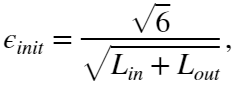
When you are done, run the code below to call randInitialWeights and initialize the neural network parameters.

initial\_Theta1 = randInitializeWeights(input\_layer\_size, hidden\_layer\_size);

initial\_Theta2 = randInitializeWeights(hidden\_layer\_size, num\_labels);

% Unroll parameters

initial\_nn\_params = [initial\_Theta1(:) ; initial\_Theta2(:)];

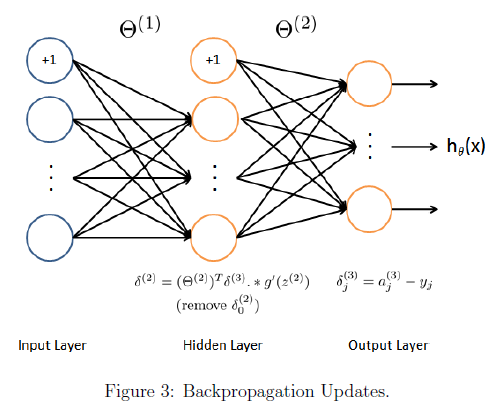
\*One effective strategy for choosing  is to base it on the number of units in the network. A good choice of  is  where  and  are the number of units in the layers adjacent to .

*You do not need to submit any code for this part of the exercise.*

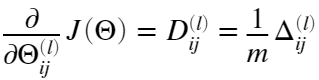
**2.3 Backpropagation**

Now, you will implement the backpropagation algorithm. Recall that the intuition behind the backpropagation algorithm is as follows. Given a training example , we will first run a 'forward pass' to compute all the activations throughout the network, including the output value of the hypothesis . Then, for each node *j* in layer *l*, we would like to compute an 'error term'  that measures how much that node was 'responsible' for any errors in our output.

For an output node, we can directly measure the difference between the network's activation and the true target value, and use that to define  (since layer 3 is the output layer). For the hidden units, you will compute  based on a weighted average of the error terms of the nodes in layer .



In detail, here is the backpropagation algorithm (also depicted in Figure 3). You should implement steps 1 to 4 in a loop that processes one example at a time. Concretely, you should implement a for loop for t = 1:m and place steps 1-4 below inside the for loop, with the  iteration performing the calculation on the  training example . Step 5 will divide the accumulated gradients by *m* to obtain the gradients for the neural network cost function.

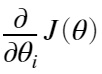
1. Set the input layer's values  to the *t*-th training example . Perform a feedforward pass (Figure 2), computing the activations  for layers 2 and 3. Note that you need to add a  term to ensure that the vectors of activations for layers  and  also include the bias unit. In MATLAB, if a\_1 is a column vector, adding one corresponds to a\_1 = [1; a\_1].
2. For each output unit *k* in layer 3 (the output layer), set  where  indicates whether the current training example belongs to class *k* , or if it belongs to a different class .You may find logical arrays helpful for this task (explained in the previous programming exercise).
3. For the hidden layer , set 
4. Accumulate the gradient from this example using the following formula: . Note that you should skip or remove . In MATLAB, removing  corresponds to delta\_2 = delta\_2(2:end).
5. Obtain the (unregularized) gradient for the neural network cost function by dividing the accumulated gradients by  

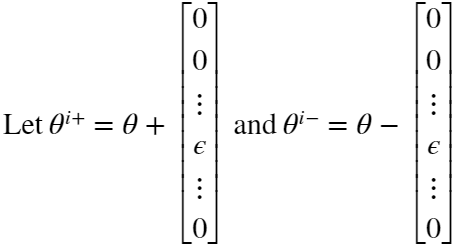
**MATLAB Tip:** You should implement the backpropagation algorithm only after you have successfully completed the feedforward and cost functions. While implementing the backpropagation algorithm, it is often useful to use the size function to print out the sizes of the variables you are working with if you run into dimension mismatch errors (''nonconformant arguments") errors.

After you have implemented the backpropagation algorithm, the code in the next section will run gradient checking on your implementation. The gradient check will allow you to increase your confidence that your code is computing the gradients correctly.

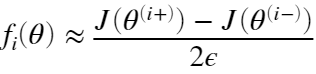
**2.4 Gradient checking**

In your neural network, you are minimizing the cost function . To perform gradient checking on your parameters, you can imagine 'unrolling' the parameters  into a long vector *θ*. By doing so, you can think of the cost function being  instead and use the following gradient checking procedure.

Suppose you have a function  that purportedly computes ; you'd like to check if  is outputting correct derivative values.



So,  is the same as *θ*, except its *i*-th element has been incremented by *ϵ*. Similarly, is the corresponding vector with the *i*-th element decreased by *ϵ*. You can now numerically verify 's correctness by checking, for each *i*, that:



The degree to which these two values should approximate each other will depend on the details of *J*. But assuming , you'll usually find that the left- and right-hand sides of the above will agree to at least 4 signicant digits (and often many more). We have implemented the function to compute the numerical gradient for you in computeNumericalGradient.m. While you are not required to modify the file, we highly encourage you to take a look at the code to understand how it works.

The code below will run the provided function checkNNGradients.m which will create a small neural network and dataset that will be used for checking your gradients. If your backpropagation implementation is correct, you should see a relative dierence that is less than 1e-9.

checkNNGradients;

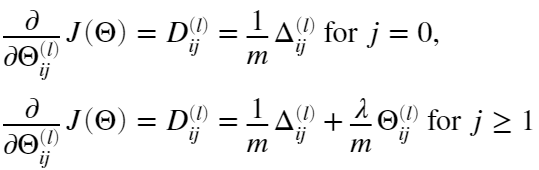
**Practical Tip:** When performing gradient checking, it is much more efficient to use a small neural network with a relatively small number of input units and hidden units, thus having a relatively small number of parameters. Each dimension of *θ* requires two evaluations of the cost function and this can be expensive. In the function checkNNGradients, our code creates a small random model and dataset which is used with computeNumericalGradient for gradient checking. Furthermore, after you are condent that your gradient computations are correct, you should turn off gradient checking before running your learning algorithm.

**Practical Tip:** Gradient checking works for any function where you are computing the cost and the gradient. Concretely, you can use the same computeNumericalGradient.m function to check if your gradient implementations for the other exercises are correct too (e.g. logistic regression's cost function). Once your cost function passes the gradient check for the (unregularized) neural network cost function, you should submit the neural network gradient function (backpropagation).

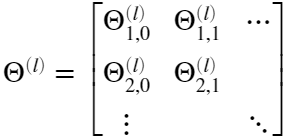
*You should now submit your solutions. Enter* **submit***at the command prompt, then enter or confirm your login and token when prompted.*

**2.5 Regularized neural networks**

After you have successfully implemeted the backpropagation algorithm, you will add regularization to the gradient. To account for regularization, it turns out that you can add this as an additional term after computing the gradients using backpropagation. Specically, after you have computed  using backpropagation, you should add regularization using



Note that you should *not* be regularizing the first column of  which is used for the bias term. Furthermore, in the parameters , *i* is indexed starting from 1, and *j* is indexed starting from 0. Thus,



Somewhat confusingly, indexing in MATLAB starts from 1 (for both *i* and *j*), thus Theta1(2, 1) actually corresponds to  (i.e., the entry in the second row, first column of the matrix  shown above)

Now, modify your code that computes grad in nnCostFunction to account for regularization. After you are done, run the code below to run gradient checking on your implementation. If your code is correct, you should expect to see a relative difference that is less than 1e-9.

% Check gradients by running checkNNGradients

lambda = 3;

checkNNGradients(lambda);

% Also output the costFunction debugging value

% This value should be about 0.576051

debug\_J = nnCostFunction(nn\_params, input\_layer\_size, hidden\_layer\_size, num\_labels, X, y, lambda);

fprintf('Cost at (fixed) debugging parameters (w/ lambda = 3): %f', debug\_J);

*You should now submit your solutions. Enter* **submit***at the command prompt, then enter or confirm your login and token when prompted.*

**2.6 Learning parameters using fmincg**

After you have successfully implemented the neural network cost function and gradient computation, run the code below to use fmincg to learn a good set of parameters. After the training completes, the code will report the training accuracy of your classifier by computing the percentage of examples it got correct. If your implementation is correct, you should see a reported training accuracy of about 95.3% (this may vary by about 1% due to the random initialization). It is possible to get higher training accuracies by training the neural network for more iterations.

options = optimset('MaxIter', 50);

lambda = 1;

% Create "short hand" for the cost function to be minimized

costFunction = @(p) nnCostFunction(p, input\_layer\_size, hidden\_layer\_size, num\_labels, X, y, lambda);

% Now, costFunction is a function that takes in only one argument (the

% neural network parameters)

[nn\_params, ~] = fmincg(costFunction, initial\_nn\_params, options);

% Obtain Theta1 and Theta2 back from nn\_params

Theta1 = reshape(nn\_params(1:hidden\_layer\_size \* (input\_layer\_size + 1)), hidden\_layer\_size, (input\_layer\_size + 1));

Theta2 = reshape(nn\_params((1 + (hidden\_layer\_size \* (input\_layer\_size + 1))):end), num\_labels, (hidden\_layer\_size + 1));

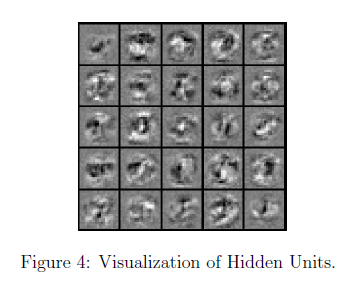
pred = predict(Theta1, Theta2, X);

fprintf('\nTraining Set Accuracy: %f\n', mean(double(pred == y)) \* 100);

**3. Visualizing the hidden layer**

One way to understand what your neural network is learning is to visualize what the representations captured by the hidden units. Informally, given a particular hidden unit, one way to visualize what it computes is to find an input x that will cause it to activate (that is, to have an activation value close to 1). For the neural network you trained, notice that the  row of  is a 401-dimensional vector that represents the parameter for the  hidden unit. If we discard the bias term, we get a 400 dimensional vector that represents the weights from each input pixel to the hidden unit.

Thus, one way to visualize the 'representation' captured by the hidden unit is to reshape this 400 dimensional vector into a 20 x 20 image and display it.\*



The code below does this by using the displayData function and it will show you an image (similar to Figure 4) with 25 units, each corresponding to one hidden unit in the network. In your trained network, you should find that the hidden units corresponds roughly to detectors that look for strokes and other patterns in the input.

% Visualize Weights

displayData(Theta1(:, 2:end));

*\*It turns out that this is equivalent to finding the input that gives the highest activation for the hidden unit, given a 'norm' constraint on the input (i.e.,* *).*

**3.1 Optional (ungraded) exercise**

In this part of the exercise, you will get to try out different learning settings for the neural network to see how the performance of the neural network varies with the regularization parameter and number of training steps (the MaxIter option when using fmincg). Neural networks are very powerful models that can form highly complex decision boundaries. Without regularization, it is possible for a neural network to 'overfit' a training set so that it obtains close to 100% accuracy on the training set but does not do as well on new examples that it has not seen before. You can set the regularization *λ* to a smaller value and the MaxIter parameter to a higher number of iterations to see this for youself. You will also be able to see for yourself the changes in the visualizations of the hidden units when you change the learning parameters *λ* and MaxIter.

% Change lambda and MaxIter to see how it affects the result

lambda = 3;

MaxIter = 50;

options = optimset('MaxIter', MaxIter);

% Create "short hand" for the cost function to be minimized

costFunction = @(p) nnCostFunction(p,input\_layer\_size, hidden\_layer\_size, num\_labels, X, y, lambda);

% Now, costFunction is a function that takes in only one argument (the neural network parameters)

[nn\_params, ~] = fmincg(costFunction, initial\_nn\_params, options);

% Obtain Theta1 and Theta2 back from nn\_params

Theta1 = reshape(nn\_params(1:hidden\_layer\_size \* (input\_layer\_size + 1)), hidden\_layer\_size, (input\_layer\_size + 1));

Theta2 = reshape(nn\_params((1 + (hidden\_layer\_size \* (input\_layer\_size + 1))):end), num\_labels, (hidden\_layer\_size + 1));

pred = predict(Theta1, Theta2, X);

fprintf('\nTraining Set Accuracy: %f\n', mean(double(pred == y)) \* 100);

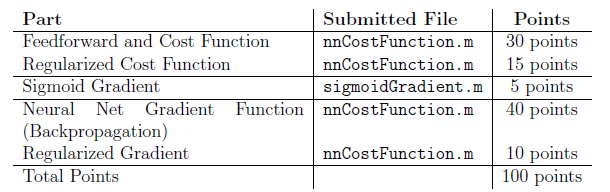
% Visualize Weights

displayData(Theta1(:, 2:end));

*You do not need to submit any solutions for this optional (ungraded) exercise.*

**Submission and Grading**

After completing various parts of the assignment, be sure to use the submit function system to submit your solutions to our servers. The following is a breakdown of how each part of this exercise is scored.



You are allowed to submit your solutions multiple times, and we will take only the highest score into consideration.