CS777 – Week 3 Homework Submission Template

**!!!! PLEASE RENAME THIS DOCUMENT WITH YOUR NAME AND LASTNAME !!!!**

**Task 1 – Simple Linear Regression**

* Calculate the m slope and the b intercept for the small dataset

|  |
| --- |
| *Console:*  *(m: 2.701426407735538 , b: 3.9501934248399837 )*  *Execution Time for Task 1: 14.96433138847351 seconds*  *Output file:*  *(2.701426407735538, 3.9501934248399837, 14.96433138847351)* |

* Calculate the m slope and the b intercept for the large dataset

|  |
| --- |
| *Console:*  *(m: 9.526356295808061e-06 , b: 12.322131336827626 )*  *Execution Time for Task 1: 434.91911458969116 seconds*  *Output file:*  *(9.526356295808061e-06, 12.322131336827626, 434.91911458969116)* |

**Task 2 – Find the Parameters using Gradient Descent**

* Print out the costs and model parameters in each iteration. The maximum number of iterations is 50.

|  |
| --- |
| *Console:*  *Iteration 1: Cost=7.735945777953018e+27, m=5593987536.256959, b=142922.24919899768*  *Iteration 2: Cost=1.3823948236008744e+53, m=-2.3647260485796726e+22, b=-7.83404841689977e+16*  *Iteration 3: Cost=2.470306156725324e+78, m=9.996320599936906e+34, b=3.3116587647501034e+29*  *Iteration 4: Cost=4.414377429511322e+103, m=-4.225708326633628e+47, b=-1.3999254903121372e+42*  *Iteration 5: Cost=7.888385833025122e+128, m=1.786318344150283e+60, b=5.917854216371133e+54*  *Iteration 6: Cost=1.409635493232824e+154, m=-7.551238703724644e+72, b=-2.501633034659049e+67*  *Iteration 7: Cost=2.5189845753110078e+179, m=3.192107730791541e+85, b=1.0575062533267967e+80*  *Iteration 8: Cost=4.501364587654818e+204, m=-1.3493881156239882e+98, b=-4.4703578036402734e+92*  *Iteration 9: Cost=8.043829783406439e+229, m=5.704219406556387e+110, b=1.889738129644882e+105*  *Iteration 10: Cost=1.4374129516696905e+255, m=-2.4113239668773693e+123, b=-7.988421409377475e+117*  *Iteration 11: Cost=2.5686222225341935e+280, m=1.0193302288664854e+136, b=3.376916389269935e+130*  *Output file:*  *(-4.3089776809742125e+148, -1.4275116090546039e+143)* |

* Calculate the m slope and the b intercept based on your calculations

|  |
| --- |
| *(m,b):*  *(-4.3089776809742125e+148, -1.4275116090546039e+143)* |

* Comment on how you can interpret the parameters of the model. What is the meaning of m and b in this case?

|  |
| --- |
| *The values of m as -4.3089776809742125e+148 and b as -1.4275116090546039e+143, it is evident that the model hasn't converged to a sensible solution and is essentially nonfunctional for predicting new values. The excessively large magnitude of these numbers indicates a severe overshooting of the optimal values due to the divergence in the gradient descent process.*  *The value of m is a large negative number, which, under normal circumstances, would indicate a steep, negatively sloped line. In practical terms, it would mean that for every unit increase in the independent variable x, the dependent variable y would decrease by 4.3089776809742125e+148 units. However, given the unrealistic magnitude of this value, it does not reflect any actual relationship between the variables and cannot be interpreted in a meaningful way.*  *The b value is also significantly large and negative. In a converged model, it would represent the value of the dependent variable y when the independent variable x is 0. However, given its unrealistic magnitude, it does not convey any meaningful information about the data in its current state*  *The values of m and b obtained are not interpretable due to their abnormal magnitude and sign, suggesting no realistic relationship between the dependent and independent variables.*  *The values are non-representative of the dataset and non-predictive, implying that any prediction or interpretation based on these values would be significantly erroneous and unreliable* |

**Task 3 – Fit Multiple Linear Regression using Gradient Descent**

* Print out the costs and model parameters in each iteration.

|  |
| --- |
| *Iteration 0, Cost: 7.379396230425838, b: 0, m1: 0, m2: 0, m3: 0, m4: 0, learning\_rate: 1e-07*  *Iteration 1, Cost: -84665197.95240544, b: 1.4758792460851675e-06, m1: 206688.14572324773, m2: 2109.3598139661317, m3: 3109.770777644441, m4: 62.73566957933406, learning\_rate: 1.05e-07*  *Iteration 2, Cost: 1019948696038598.1, b: -17.7796900941259, m1: -2489940275829.8667, m2: -25411132983.13612, m3: -37462929868.3806, m4: -755766954.4623939, learning\_rate: 1.1025000000000001e-07*  *Iteration 3, Cost: -1.2901524361995425e+22, b: 224898669.69682083, m1: 3.1495726454966796e+19, m2: 3.2143023714932096e+17, m3: 4.7387570006789286e+17, m4: 9559839443746302.0, learning\_rate: 5.5125000000000006e-08*  *Iteration 4, Cost: 8.159690695754973e+28, b: -1422392836011326.2, m1: -1.9919769083076286e+26, m2: -2.032915833672798e+24, m3: -2.997071533793741e+24, m4: -6.046210569646363e+22, learning\_rate: 5.788125000000001e-08*  *Iteration 5, Cost: -5.418707001414941e+35, b: 9.445860519280514e+21, m1: 1.3228368111206984e+33, m2: 1.3500236310355836e+31, m3: 1.9903024648175585e+31, m4: 4.015182041497878e+29, learning\_rate: 2.8940625000000004e-08*  *Iteration 6, Cost: 1.799233711209285e+42, b: -3.136414401670434e+28, m1: -4.3923625772263403e+39, m2: -4.482634007068669e+37, m3: -6.608623218172857e+37, m4: -1.3332056676654864e+36, learning\_rate: 3.0387656250000004e-08*  *Iteration 7, Cost: -6.272906235361222e+48, b: 1.0934895969513505e+35, m1: 1.5313674053012186e+46, m2: 1.5628399267199245e+44, m3: 2.3040516378812722e+44, m4: 4.648131086926362e+42, learning\_rate: 1.5193828125000002e-08*  *Iteration 8, Cost: 1.0935027813802206e+55, b: -1.906188090196787e+41, m1: -2.6695035031326496e+52, m2: -2.724366892472667e+50, m3: -4.016458687464843e+50, m4: -8.102694478552656e+48, learning\_rate: 1.5953519531250003e-08*  *Iteration 9, Cost: -2.0015216818485225e+61, b: 3.48904168985702e+47, m1: 4.8861962056890374e+58, m2: 4.986616858634371e+56, m3: 7.3516311838390184e+56, m4: 1.4830980731248804e+55, learning\_rate: 7.976759765625001e-09*  *Iteration 10, Cost: 1.831768022200894e+67, b: -3.193128035317386e+53, m1: -4.471786661593739e+64, m2: -4.563690407879373e+62, m3: -6.728122425900822e+62, m4: -1.3573131127058325e+61, learning\_rate: 8.375597753906251e-09*  *Iteration 11, Cost: -1.7602322306986412e+73, b: 3.0684272333565852e+59, m1: 4.2971505753702434e+70, m2: 4.38546521694781e+68, m3: 6.465369960944658e+68, m4: 1.3043061453076388e+67, learning\_rate: 4.187798876953126e-09*  *Iteration 12, Cost: 8.457441783295334e+78, b: -1.4742966433520597e+65, m1: -2.064665115853667e+76, m2: -2.1070979225440888e+74, m3: -3.1064361337397357e+74, m4: -6.26683973804716e+72, learning\_rate: 4.397188820800782e-09*  *Iteration 13, Cost: -4.266752064154851e+84, b: 7.4377789494495e+70, m1: 1.0416168825727974e+82, m2: 1.0630240965002807e+80, m3: 1.5671869964246583e+80, m4: 3.1616004074486696e+78, learning\_rate: 2.198594410400391e-09*  *Iteration 14, Cost: 1.0762791845117789e+90, b: -1.8761640099840875e+76, m1: -2.6274565573363073e+87, m2: -2.6814558017313657e+85, m3: -3.953196054347149e+85, m4: -7.975070163715003e+83, learning\_rate: 2.3085241309204107e-09*  *Iteration 15, Cost: -2.850636593096208e+95, b: 4.969214176465466e+81, m1: 6.959090091954947e+92, m2: 7.102112668519894e+90, m3: 1.0470448090396856e+91, m4: 2.1122797103530124e+89, learning\_rate: 1.1542620654602053e-09*  *Iteration 16, Cost: 3.775088481701438e+100, b: -6.58071367150558e+86, m1: -9.215899674089083e+97, m2: -9.405303992661714e+95, m3: -1.3865979297409526e+96, m4: -2.797284937686267e+94, learning\_rate: 1.2119751687332157e-09*  *Iteration 17, Cost: -5.249305756423569e+105, b: 9.150561192049123e+91, m1: 1.2814818896115867e+103, m2: 1.3078187870007222e+101, m3: 1.9280810316672443e+101, m4: 3.8896582151460355e+99, learning\_rate: 6.059875843666078e-10*  *Iteration 18, Cost: 3.649585021920643e+110, b: -6.361936724261774e+96, m1: -8.909515519201351e+107, m2: -9.092623058931964e+105, m3: -1.340500245315503e+106, m4: -2.704288723326121e+104, learning\_rate: 6.362869635849383e-10*  *Iteration 19, Cost: -2.664248302857507e+115, b: 4.6443031245186694e+101, m1: 6.504071410514679e+112, m2: 6.63774248518141e+110, m3: 9.785839984850289e+110, m4: 1.974168733783106e+109, learning\_rate: 3.1814348179246915e-10*  *Iteration 20, Cost: 9.724560111134022e+119, b: -1.6951800198302832e+106, m1: -2.3739992001052962e+117, m2: -2.422789412313456e+115, m3: -3.571851357418391e+115, m4: -7.205755747542005e+113, learning\_rate: 3.340506558820926e-10*  *Iteration 21, Cost: -3.726963146022393e+124, b: 6.496821848576327e+110, m1: 9.098414145590842e+121, m2: 9.285403912437187e+119, m3: 1.3689213928480653e+120, m4: 2.7616247730917864e+118, learning\_rate: 1.670253279410463e-10*  *Iteration 22, Cost: 7.141655613364473e+128, b: -1.244929515158682e+115, m1: -1.7434500398784064e+126, m2: -1.779281264006965e+124, m3: -2.6231451094232165e+124, m4: -5.291861574700175e+122, learning\_rate: 1.7537659433809863e-10*  *Iteration 23, Cost: -1.4369216471993159e+133, b: 2.5048339858633366e+119, m1: 3.5078716179249607e+130, m2: 3.579965071296784e+128, m3: 5.277843395894566e+128, m4: 1.0647377670296237e+127, learning\_rate: 8.768829716904931e-11*  *Iteration 24, Cost: 1.4454919923126858e+137, b: -2.5197737647664828e+123, m1: -3.5287938932885938e+134, m2: -3.601317339330518e+132, m3: -5.309322453534754e+132, m4: -1.0710882664715138e+131, learning\_rate: 9.207271202750178e-11*  *Iteration 25, Cost: -1.5268263543812998e+141, b: 2.6615553815488384e+127, m1: 3.727350648850547e+138, m2: 3.8039548149865645e+136, m3: 5.608065274023951e+136, m4: 1.1313558302744154e+135, learning\_rate: 4.603635601375089e-11*  *Iteration 26, Cost: 8.062922614942519e+144, b: -1.405524276891343e+131, m1: -1.9683534905057962e+142, m2: -2.008806909571055e+140, m3: -2.961528414428497e+140, m4: -5.97450684767603e+138, learning\_rate: 4.8338173814438435e-11*  *Iteration 27, Cost: -4.470834059651482e+148, b: 7.793533571992269e+134, m1: 1.0914382100702066e+146, m2: 1.1138693473170336e+144, m3: 1.642146742090164e+144, m4: 3.3128221588918494e+142, learning\_rate: 2.4169086907219217e-11*  *Iteration 28, Cost: 1.2392995284982557e+152, b: -2.1603401853522484e+138, m1: -3.025428457147006e+149, m2: -3.0876067832552295e+147, m3: -4.551973202414499e+147, m4: -9.183026891035704e+145, learning\_rate: 2.5377541252580178e-11*  *Iteration 29, Cost: -3.6071215558412246e+155, b: 6.287914641568178e+141, m1: 8.805851977247664e+152, m2: 8.986829033442334e+150, m3: 1.3249033250205995e+151, m4: 2.672823920675379e+149, learning\_rate: 1.2688770626290089e-11*  *Iteration 30, Cost: 5.247664231072361e+158, b: -9.14769969400162e+144, m1: -1.2810811537607385e+156, m2: -1.3074098152636943e+154, m3: -1.927478095957428e+154, m4: -3.888441869049387e+152, learning\_rate: 1.3323209157604595e-11*  *Iteration 31, Cost: -8.01631574558152e+161, b: 1.3973997928197468e+148, m1: 1.9569756318349702e+159, m2: 1.997195214200116e+157, m3: 2.944409613404427e+157, m4: 5.939971844248738e+155, learning\_rate: 6.661604578802297e-12*  *Iteration 32, Cost: 6.118841485627598e+164, b: -1.0666331137249963e+151, m1: -1.4937564914447718e+162, m2: -1.5244560368370678e+160, m3: -2.247463331656933e+160, m4: -4.533971377572047e+158, learning\_rate: 6.994684807742413e-12*  *Iteration 33, Cost: -4.904333356940134e+167, b: 8.549207184963426e+153, m1: 1.1972658231701897e+165, m2: 1.2218719197432042e+163, m3: 1.8013719446457325e+163, m4: 3.634038750418094e+161, learning\_rate: 3.4973424038712064e-12*  *Iteration 34, Cost: 1.9629922505767818e+170, b: -3.4218773952043874e+156, m1: -4.792136589650765e+167, m2: -4.890624138058185e+165, m3: -7.210111773381908e+165, m4: -1.45454833229728e+164, learning\_rate: 3.672209524064767e-12*  *Iteration 35, Cost: -8.250839754415065e+172, b: 1.4382818902514735e+159, m1: 2.0142285875483478e+170, m2: 2.0556248273692656e+168, m3: 3.030550774518538e+168, m4: 6.1137506790003e+166, learning\_rate: 1.8361047620323834e-12*  *Iteration 36, Cost: 1.7298691140551733e+175, b: -3.015498413867005e+161, m1: -4.223026899028565e+172, m2: -4.309818157658825e+170, m3: -6.353845595668493e+170, m4: -1.281807523286018e+169, learning\_rate: 1.9279100001340025e-12*  *Iteration 37, Cost: -3.809046408047599e+177, b: 6.639908943681161e+163, m1: 9.298799146210586e+174, m2: 9.489907207074525e+172, m3: 1.3990707474240134e+173, m4: 2.822447260726697e+171, learning\_rate: 9.639550000670012e-13*  *Iteration 38, Cost: 4.174576823170629e+179, b: -7.2770995716126565e+165, m1: -1.0191146875259092e+177, m2: -1.0400594384199434e+175, m3: -1.533330836776603e+175, m4: -3.0932999121139485e+173, learning\_rate: 1.0121527500703514e-12*  *Iteration 39, Cost: -4.8060314452801516e+181, b: 8.377847828188079e+167, m1: 1.1732679603381373e+179, m2: 1.1973808550515783e+177, m3: 1.7652654459881064e+177, m4: 3.5611984823914176e+175, learning\_rate: 5.060763750351757e-13*  *Iteration 40, Cost: 2.7424704045573946e+183, b: -4.7806594659830096e+169, m1: -6.695030389371707e+180, m2: -6.83262603532771e+178, m3: -1.0073151407622232e+179, m4: -2.03213015851243e+177, learning\_rate: 5.313801937869346e-13*  *Iteration 41, Cost: -1.6445567211444914e+185, b: 2.866782315397452e+171, m1: 4.014758812642762e+182, m2: 4.0972697648051426e+180, m3: 6.040491384331324e+180, m4: 1.2185922972472934e+179, learning\_rate: 2.656900968934673e-13*  *Iteration 42, Cost: 4.848667324319208e+186, b: -8.45217046021391e+172, m1: -1.1836764046871159e+184, m2: -1.2080032127870993e+182, m3: -1.7809256939253696e+182, m4: -3.592791040504873e+180, learning\_rate: 2.789746017381407e-13*  *Iteration 43, Cost: -1.5034399718361532e+188, b: 2.620788366923236e+174, m1: 3.670258859789966e+185, m2: 3.7456896807523687e+183, m3: 5.522166599650162e+183, m4: 1.1140268654145078e+182, learning\_rate: 1.3948730086907034e-13*  *Iteration 44, Cost: 2.255707393336596e+189, b: -3.932136837109599e+175, m1: -5.50672471171371e+186, m2: -5.619898409179367e+184, m3: -8.285260641870313e+184, m4: -1.6714459398218287e+183, learning\_rate: 1.4646166591252386e-13*  *Iteration 45, Cost: -3.564880101814096e+190, b: 6.214279569074532e+176, m1: 8.702730420153622e+187, m2: 8.881588131635102e+185, m3: 1.3093879502189094e+186, m4: 2.6415236256840016e+184, learning\_rate: 7.323083295626193e-14*  *Iteration 46, Cost: 2.6386927522764855e+191, b: -4.599754827993548e+177, m1: -6.44168416014626e+188, m2: -6.574073057807865e+186, m3: -9.691973910715069e+186, m4: -1.9552324473725782e+185, learning\_rate: 7.689237460407504e-14*  *Iteration 47, Cost: -2.063987376645621e+192, b: 3.597931548662672e+178, m1: 5.038690002619274e+189, m2: 5.14224469398902e+187, m3: 7.581068993059423e+187, m4: 1.5293842325155756e+186, learning\_rate: 3.844618730203752e-14*  *Iteration 48, Cost: 7.040268395389643e+192, b: -1.2272557505649048e+179, m1: -1.7186989795086197e+190, m2: -1.7540215221314027e+188, m3: -2.5859053712735798e+188, m4: -5.216735140154855e+186, learning\_rate: 4.03684966671394e-14*  *Iteration 49, Cost: -2.5567113386923707e+193, b: 4.456845274536167e+179, m1: 6.241547796086712e+190, m2: 6.369823509686336e+188, m3: 9.390854456419054e+188, m4: 1.8944854279317004e+187, learning\_rate: 2.01842483335697e-14* |

* What are m and b values based on your calculations?

|  |
| --- |
| *(b, m1, m2, m3, m4):*  *(-5.86421404094788e+179, -8.212484384903444e+190, -8.381266605166138e+188, -1.2356269326751942e+189, -2.492720155739405e+187)* |

* Comment on how you can interpret the parameters of the model. What is the meaning of *mi* and *b* in this case

|  |
| --- |
| b has a value of −5.86421404094788e+179. If all independent variables are zero, theoretically, the total\_amount would be this value, although practically this value is non-sensical as it’s exceptionally high and negative. In the context of this task, a possible explanation for this unrealistic value could be that the model has not been properly trained or converged due to learning rate issues, incorrect data scaling, or other problems in the training process.  m1 has a value of −8.212484384903444e+190. This value suggests that for each additional second in trip time, the total amount paid decreases by this value, holding other variables constant. This is counterintuitive as we would generally expect the total amount to increase with the trip time. This again suggests a possible issue with model convergence or training.  m2 value is −8.381266605166138e+188. Theoretically, for each additional mile in trip distance, the total amount decreases by this value, which is illogical as we generally expect the total amount to increase with distance.  value of m3 is−1.2356269326751942e+189, it implies that for each additional dollar in fare amount, the total amount decreases by this value, holding other factors constant. This is not logical as the total amount should increase with the fare amount.  m4 is −2.492720155739405e+187. This means, for each additional dollar in tolls amount, the total amount decreases by this value, which is again counterintuitive as we would expect the total amount to increase with the toll amount.  The parameters obtained, both the intercept and the coefficients, are not logically coherent with real-world expectations. They indicate potential issues in model training, and they could be a result of non-convergence of the algorithm, numerical instability, inappropriate learning rate, data scaling issues, or even outliers in the data affecting the model. Further examination and adjustments in the model training process are essential to obtain meaningful and accurate parameters. |

**Spark History Output:**

**Task 1:**

|  |
| --- |
| *spark-history* |

**Task 2:**

|  |
| --- |
| *spark-history* |

**Task 3:**

|  |
| --- |
| *spark-history* |