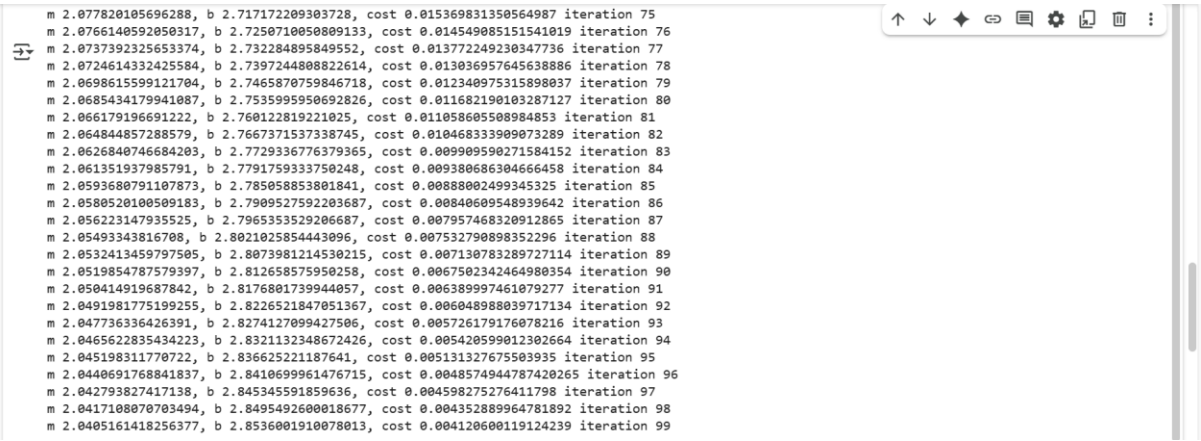
Activity 4

In this activity, we explored how the gradient descent algorithm minimises the cost function during model training. By running a tutorial, we visualised how the cost value decreases over iterations. We also experimented with different learning rates as part of hyperparameter optimization to see how they affect the convergence speed and stability. This helped us understand the critical role hyperparameters play in achieving optimal model performance.

1.

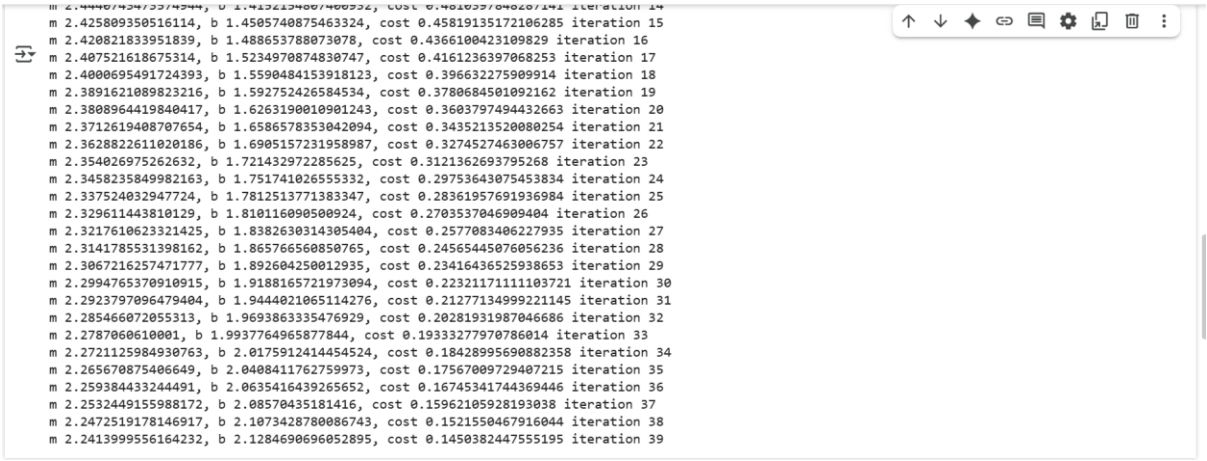
100 Iterations and Learning rate is 0.08



```
m 2.077820105696288, b 2.717172209303728, cost 0.015369831350564987 iteration 75
m 2.0766140592050317, b 2.7250710050809133, cost 0.014549085151541019 iteration 76
m 2.0737392325653374, b 2.732284895849552, cost 0.013772249230347736 iteration 77
m 2.0724614332425584, b 2.7397244808822614, cost 0.013036957645638886 iteration 78
m 2.0698615599121704, b 2.7465870759846718, cost 0.012340975315898037 iteration 79
m 2.0685434179941087, b 2.7535995950692826, cost 0.011682190103287127 iteration 80
m 2.066179196691222, b 2.760122819221025, cost 0.011058605508984853 iteration 81
m 2.064844857288579, b 2.7667371537338745, cost 0.010468333909073289 iteration 82
m 2.0626840746684203, b 2.7729336776379365, cost 0.009909590271584152 iteration 83
m 2.061351937985791, b 2.7791759333750248, cost 0.009380686304666458 iteration 84
m 2.0593680791107873, b 2.785058853801841, cost 0.00888002499345325 iteration 85
m 2.0580520100509183, b 2.7909527592203687, cost 0.00840609548939642 iteration 86
m 2.056223147935525, b 2.7965353529206687, cost 0.007957468320912865 iteration 87
m 2.05493343816708, b 2.8021025854443096, cost 0.007532790898352296 iteration 88
m 2.0532413459797505, b 2.8073981214530215, cost 0.007130783289727114 iteration 89
m 2.0519854787579397, b 2.812658575950258, cost 0.006750234246498035 4 iteration 90
m 2.050414919687842, b 2.8176801739944057, cost 0.006389997461079277 iteration 91
m 2.0491981775199255, b 2.8226521847051367, cost 0.006048988039717134 iteration 92
m 2.047736336426391, b 2.8274127099427506, cost 0.005726179176078216 iteration 93
m 2.0465622835434223, b 2.8321132348672426, cost 0.005420599012302664 iteration 94
m 2.045198311770722, b 2.836625221187641, cost 0.005131327675503935 iteration 95
m 2.0440691768841837, b 2.8410699961476715, cost 0.004857494478742026 5 iteration 96
m 2.042793827417138, b 2.845345591859636, cost 0.004598275276411798 iteration 97
m 2.0417108070703494, b 2.8495492600018677, cost 0.004352889964781892 iteration 98
m 2.0405161418256377, b 2.8536001910078013, cost 0.004120600119124239 iteration 99
```

2.

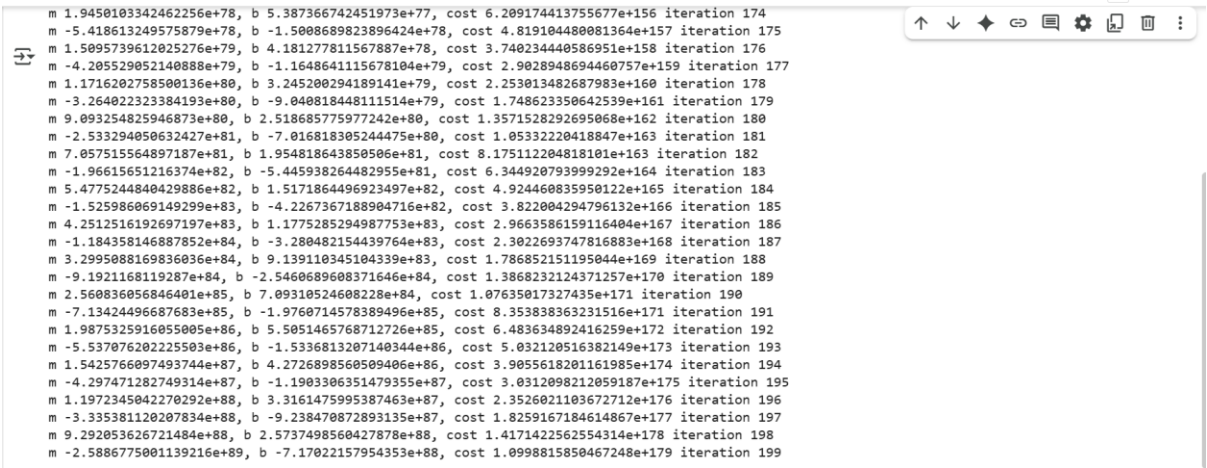
40 Iterations and Learning rate is 0.07



```
m 2.425809350516114, b 1.4505740875463324, cost 0.45819135172106285 iteration 15
m 2.4208218339951839, b 1.488653788073078, cost 0.4366100423109829 iteration 16
m 2.407521618675314, b 1.5234970874830747, cost 0.4161236397068253 iteration 17
m 2.4000695491724393, b 1.5590484153918123, cost 0.396632275909914 iteration 18
m 2.3891621089823216, b 1.592752426584534, cost 0.3780684501092162 iteration 19
m 2.3808964419840417, b 1.6263190010901243, cost 0.3603797494432663 iteration 20
m 2.3712619408707654, b 1.6586578353042094, cost 0.3435213520080254 iteration 21
m 2.3628822611020186, b 1.6905157231958987, cost 0.3274527463006757 iteration 22
m 2.354026975262632, b 1.721432972285625, cost 0.3121362693795268 iteration 23
m 2.3458235849982163, b 1.751741026555332, cost 0.29753643075453834 iteration 24
m 2.337524032947724, b 1.7812513771383347, cost 0.28361957691936984 iteration 25
m 2.329611443810129, b 1.810116090500924, cost 0.2703537046909404 iteration 26
m 2.3217610623321425, b 1.8382630314305404, cost 0.2577083406227935 iteration 27
m 2.3141785531398162, b 1.865766560850765, cost 0.24565445076056236 iteration 28
m 2.3067216257471777, b 1.892604250012935, cost 0.23416436525938653 iteration 29
m 2.2994765370910915, b 1.9188165721973094, cost 0.22321171111103721 iteration 30
m 2.2923797096479404, b 1.9444021065114276, cost 0.21277134999221145 iteration 31
m 2.285466072055313, b 1.9693863335476929, cost 0.20281931987046686 iteration 32
m 2.2787060610001, b 1.9937764965877844, cost 0.19333277970786014 iteration 33
m 2.2721125984930763, b 2.0175912414454524, cost 0.18428995690882358 iteration 34
m 2.265670875406649, b 2.0408411762759973, cost 0.17567009729407215 iteration 35
m 2.259384433244491, b 2.0635416439265652, cost 0.16745341744369446 iteration 36
m 2.2532449155988172, b 2.08570435181416, cost 0.15962105928193038 iteration 37
m 2.2472519178146917, b 2.1073428780086743, cost 0.1521550467916044 iteration 38
m 2.2413999556164232, b 2.1284690696052895, cost 0.1450382447555195 iteration 39
```

3.

200 Iterations and Learning rate is 0.16



```
m 1.9450103342462256e+78, b 5.387366742451973e+77, cost 6.209174413755677e+156 iteration 174
m -5.418613249575879e+78, b -1.5008689823896424e+78, cost 4.819104480081364e+157 iteration 175
m 1.5095739612025276e+79, b 4.1812778111567887e+78, cost 3.740234440586951e+158 iteration 176
m -4.205529052140888e+79, b -1.1648641115678104e+79, cost 2.9028948694460757e+159 iteration 177
m 1.1716202758500136e+80, b 3.245200294189141e+79, cost 2.253013482687983e+160 iteration 178
m -3.264022323384193e+80, b -9.040818448111514e+79, cost 1.748623350642539e+161 iteration 179
m 9.093254825946873e+80, b 2.518685775977242e+80, cost 1.3571528292695068e+162 iteration 180
m -2.533294050632427e+81, b -7.016818305244475e+80, cost 1.05332220418847e+163 iteration 181
m 7.057515564897187e+81, b 1.954818643850506e+81, cost 8.175112204818101e+163 iteration 182
m -1.966156512163746+82, b -5.445938264482955e+81, cost 6.344920793999292e+164 iteration 183
m 5.4775244840429886e+82, b 1.5171864496923497e+82, cost 4.924460835950122e+165 iteration 184
m -1.525986069149299e+83, b -4.2267367188904716e+82, cost 3.822004294796132e+166 iteration 185
m 4.2512516192697197e+83, b 1.1775285294987753e+83, cost 2.9663586159116404e+167 iteration 186
m -1.184358146887852e+84, b -3.280482154439764e+83, cost 2.3022693747816883e+168 iteration 187
m 3.2995088169836036e+84, b 9.139110345104339e+83, cost 1.786852151195044e+169 iteration 188
m -9.19211681192876+84, b -2.5460689608371646e+84, cost 1.3868232124371257e+170 iteration 189
m 2.560836056846401e+85, b 7.09310524608228e+84, cost 1.07635017327435e+171 iteration 190
m -7.134244966876836+85, b -1.9760714578389496e+85, cost 8.353838363231516e+171 iteration 191
m 1.9875325916055005e+86, b 5.505146576871272 6e+85, cost 6.483634892416259e+172 iteration 192
m -5.537076202225503e+86, b -1.5336813207140344e+86, cost 5.032120516382149e+173 iteration 193
m 1.5425766097493744e+87, b 4.272689856050940 6e+86, cost 3.9055618201161985e+174 iteration 194
m -4.297471282749314e+87, b -1.1903306351479355e+87, cost 3.0312098212059187e+175 iteration 195
m 1.1972345042270292e+88, b 3.3161475995387463e+87, cost 2.3526021103672712e+176 iteration 196
m -3.335381120207834e+88, b -9.238470872893135e+87, cost 1.8259167184614867e+177 iteration 197
m 9.292053626721484e+88, b 2.5737498560427878e+88, cost 1.4171422562554314e+178 iteration 198
m -2.5886775001139216e+89, b -7.17022157954353e+88, cost 1.0998815850467248e+179 iteration 199
```

Through this activity, we learned that simply doubling the learning rate or increasing the number of iterations doesn't guarantee better gradient descent performance. In some cases, a higher learning rate can cause the algorithm to diverge, while excessive iterations may waste resources if convergence has already been achieved. The most effective strategy is to monitor the convergence behavior and adjust the iterations only when the cost function hasn’t stabilised, ensuring both efficiency and accuracy.