1a)

$$X = \begin{bmatrix} 1 & 60 \\ 1 & 70 \\ 1 & 62 \\ 1 & 72 \\ 1 & 65 \end{bmatrix} \qquad X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 60 & 70 & 62 & 72 & 65 \end{bmatrix} \begin{bmatrix} 1 & 60 \\ 1 & 70 \\ 1 & 62 \\ 1 & 72 \\ 1 & 65 \end{bmatrix} = \begin{bmatrix} 5 & 329 \\ 329 & 21753 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{(5)(21753)-(329)(329)} \begin{bmatrix} 21753 & -329 \\ -329 & 5 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{524} \begin{bmatrix} 21753 & -329 \\ -329 & 5 \end{bmatrix}$$

$$y = \begin{bmatrix} 130 \\ 155 \\ 125 \\ 162 \\ 150 \end{bmatrix} \qquad X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 60 & 70 & 62 & 72 & 65 \end{bmatrix} \begin{bmatrix} 130 \\ 155 \\ 125 \\ 162 \\ 150 \end{bmatrix} = \begin{bmatrix} 722 \\ 47814 \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{524} \begin{bmatrix} 21753 & -329 \\ -329 & 5 \end{bmatrix} \begin{bmatrix} 722 \\ 47814 \end{bmatrix} = \begin{bmatrix} -\frac{6285}{131} \\ \frac{383}{131} \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} -47.977 \\ 2.924 \end{bmatrix}$$

Weight = $-47.977 + 2.924(\text{Height})$

1b) Predicted Weights

Student 1: $-47.977 + 2.924(60) = 127.463$

Student 2: $-47.977 + 2.924(70) = 156.703$

Student 3: $-47.977 + 2.924(62) = 133.311$

Student 4: $-47.977 + 2.924(72) = 162.551$

Student 5: $-47.977 + 2.924(65) = 142.083$

Problem 1.2a)

Closed Form Without Normalization
Beta:

```
[-0.0862246   0.05340575  0.65803045  0.41731923 -0.01772481  0.30069864
  1.02871152  0.48383363  0.26685697  0.04573456  0.31944742  1.14776959
  0.29366213  0.41491543  0.85180482 -0.05950309  0.47235562  0.46198106
  0.00497427  0.0205398   0.41310473  0.98508025  0.15573467  0.8618602
  0.41974331 -0.06893699  0.33317496  0.27766637 -0.04184791 -0.23599504
  0.15020297  0.37745027  0.80256455  0.16053288  0.2744667   0.63461071
  0.74135259  0.56079776  0.94058723 -0.0432542   0.80803615  0.93967722
  0.12225161 -0.19933624  0.09398732  0.11412993  0.35479619  0.78582876
  0.38900433  0.11804526  0.67618837  0.70377377  0.05526258 -0.24919095
  0.87339793 -0.01381723  0.83138416  0.90569236  0.39980648  0.25235308
  0.69692397 -0.00949757  0.17676599  0.45822485  0.02743899  1.16718165
  0.04176352  1.01993881  0.56015024 -0.29761224  0.3177761   0.55781578
  1.1376088   0.55190283  0.4099807   0.91987238  1.34076835  0.53297825
  0.63648277  0.22140583  0.21469531 -0.00609269  0.82898663  0.46891532
 -0.25571565  0.1972989   1.38639797  0.87219453  0.65782257  0.54983464
  1.11698567  0.94267463  0.79030138  0.30055848  0.53288973  0.22873689
  0.86702876  0.98591924  0.08132528  0.30834368  0.70121488]
```

MSE: 4.39609786082

Batch Gradient Without Normalization
Beta:

```
[ 0.0874417   0.18788154  0.25328825  0.43203334  0.40008772  0.68357969
  0.14549468  0.38820946  0.41427995  0.58190915  0.70833928  0.37018394
  0.20009263  0.65607924  0.52564605  0.01332235  0.00953677  0.16055037
  0.97544956  0.21064176  0.31507318  0.0911416   0.86469478  0.1294579
  0.31019429  0.56828753  0.53037137  0.79408013  0.70291961  0.75845567
  0.4654959   0.46493752  0.59539067  0.4274803   0.84180843  0.03492734
  0.45110928  0.7196433   0.47139916  0.74231586  0.06049337  0.348746
  0.04788297  0.05277103  0.71922795  0.6969364   0.1243031   0.10282148
  0.17618983  0.08749948  0.9879209   0.43319112  0.55862622  0.35100048
  0.70185801  0.85016305  0.5546261   0.88660503  0.25572919  0.60900557
  0.46385524  0.49501469  0.12314057  0.39197463  0.28219459  0.11170492
  0.69940394  0.50106581  0.39816276  0.14670791  0.60904015  0.51006796
  0.3249821   0.35466014  0.7205346   0.52964999  0.71688617  0.63264424
  0.94886097  0.91878499  0.98322242  0.45065191  0.71118058  0.66343054
  0.45934477  0.49747213  0.51572819  0.29873251  0.15632151  0.03201158
  0.10994919  0.41971494  0.1096648   0.00794452  0.1019796   0.17861654
  0.26555713  0.62868776  0.1894415   0.46663116  0.0515934 ]
```

MSE: 5.50765097196

Stochastic Gradient Without Normalization
Beta:

```
[-0.01949056  0.08498841  0.68500618  0.39864964 -0.00622769  0.26030889
  1.04980347  0.49091459  0.26825054  0.01738727  0.32640806  1.18653489
  0.29813395  0.37411836  0.9100955  -0.05062397  0.41118815  0.43512388
 -0.00627129 -0.00464422  0.42769804  0.93484818  0.16427188  0.90358117
  0.39196861 -0.07388137  0.33956675  0.27790862  0.01881521 -0.23005581
  0.09438022  0.39790514  0.85141525  0.17998305  0.2288911   0.63032757
  0.74928875  0.55426605  0.96746864 -0.02554746  0.82313886  1.000581
  0.11776229 -0.19419136  0.1153564   0.18571286  0.29820378  0.83283986
```

0.32176682 0.12671248 0.63542624 0.7013822   0.07767014 -0.29203656
0.91574022 -0.00956545 0.79933424 0.84052368 0.39424867 0.21544763
0.66081596 0.05631196 0.15124615 0.44396957 0.00282558 1.17883076
-0.02623702 1.04794603 0.55591121 -0.32636215 0.32037398 0.53458027
1.19453272 0.56561885 0.41085296 0.93656607 1.37215879 0.48052825
0.67643275 0.23943924 0.22161636 0.02865406 0.83583517 0.47178466
-0.28763387 0.17116754 1.40985834 0.89854527 0.6520714   0.58949091
1.05578183 0.91671195 0.79402767 0.33098161 0.58769391 0.22344091
0.84143429 1.00232216 0.07851985 0.27401424 0.6788708]
MSE: 4.46217554874

Closed Form With Normalization
Beta:
[2.27729720e+01   1.53191032e-01   1.85307313e-01   1.19941086e-01
-5.02643450e-03   8.91409482e-02   2.85334735e-01   1.40179587e-01
7.57622607e-02   1.29588244e-02   9.39644822e-02   3.31336158e-01
8.47980841e-02   1.19938006e-01   2.41980006e-01 -1.70818954e-02
1.37050979e-01   1.35282526e-01   1.41548177e-03   5.96124757e-03
1.15772937e-01   2.84695298e-01   4.40028065e-02   2.49061009e-01
1.20225794e-01 -1.97867203e-02   9.78450167e-02   8.05000257e-02
-1.21180475e-02 -6.76721206e-02   4.42719117e-02   1.07760749e-01
2.27868151e-01   4.71918067e-02   7.98329570e-02   1.82865595e-01
2.10504374e-01   1.61998603e-01   2.74318321e-01 -1.24393879e-02
2.32229995e-01   2.68686623e-01   3.49570586e-02 -5.72887604e-02
2.74420886e-02   3.22205699e-02   1.03168217e-01   2.23680975e-01
1.12389162e-01   3.34056315e-02   1.96513521e-01   2.04069259e-01
1.61178878e-02 -7.11959973e-02   2.51631165e-01 -3.88541393e-03
2.30940122e-01   2.65349086e-01   1.14181954e-01   7.19159231e-02
2.03124339e-01 -2.77783657e-03   5.09788755e-02   1.31412782e-01
7.74235920e-03   3.36612770e-01   1.19459050e-02   2.97996188e-01
1.64171823e-01 -8.56897338e-02   9.04358074e-02   1.57799695e-01
3.30413482e-01   1.58063366e-01   1.17460867e-01   2.66470115e-01
3.90423742e-01   1.54496507e-01   1.82139546e-01   6.24852555e-02
6.11567085e-02 -1.74258210e-03   2.34243793e-01   1.35090828e-01
-7.34511846e-02   5.72585257e-02   4.02764875e-01   2.50516976e-01
1.87479158e-01   1.57776498e-01   3.18066564e-01   2.66011306e-01
2.29796702e-01   8.52799113e-02   1.56628433e-01   6.56759268e-02
2.52655201e-01   2.89923735e-01   2.33168105e-02   9.00779177e-02
2.03896451e-01]
MSE: 4.40454594906

Batch Gradient With Normalization
Beta:
[2.32068671 0.7794666   0.26668133 0.31209472 0.87709819 0.2493291
0.69844256 0.85869755 0.79143572 0.87676246 0.70327194 0.91514776
0.21130097 0.75908527 0.87472337 0.52330978 0.10198293 0.51006477
0.79424782 0.73281839 0.27654786 0.09013014 0.8248391   0.53714447
0.01408347 0.18152578 0.70639232 0.39839558 0.49939316 0.49304943
-0.0242745   0.78106371 0.9312846   0.00466373 0.73612658 0.7785684
0.46913455 0.13791285 0.50624409 0.80432271 0.16464894 0.32446936
0.53754978 0.28716326 0.88945348 0.10535268 0.38902557 0.35566161
0.49140318 0.48032163 0.09118423 0.53177736 0.80172904 0.24422292
0.62657629 0.53406056 0.79887059 0.79441561 0.91687916 0.799978
0.53836416 0.56818554 0.41804427 0.35512013 0.60285292 0.69574105

```
 0.61010235  0.33718133  0.82247304  0.11727345  0.83546069  0.16648586
 0.37617875  0.81740852  0.80373827  0.02849832  0.67940167  0.70139853
 0.90316629  0.68121258  0.73549678  0.11680725  0.45256542  0.04468106
 0.20410037  0.85919305  0.67425869  0.03304936  0.35717255  0.81454884
 0.46193562  0.60836914  0.92393604  0.18601277  0.06963185  0.42989481
 0.31171928  0.33068151  0.75184304  0.75554609  0.56851976]
MSE:  440.992321592
-----------------------------------------------
Stochastic Gradient With Normalization
Beta:
[ 2.28017847e+01   1.61736736e-01   1.77489581e-01   1.54000306e-01
   5.34916623e-03   7.73686749e-02   2.98385677e-01   1.18370739e-01
   9.25267981e-02   1.32965238e-02   8.68603632e-02   3.18072514e-01
   7.19071560e-02   1.48432233e-01   2.52481561e-01  -2.67710346e-02
   1.41958286e-01   1.00082204e-01  -1.24210011e-02   1.94223490e-02
   1.14430336e-01   2.93412647e-01   4.37205938e-02   2.80213971e-01
   1.21716761e-01  -1.96747525e-02   5.61665389e-02   8.99042231e-02
  -4.02324374e-03  -8.77438895e-02   4.97268862e-02   1.30418091e-01
   2.25242689e-01   2.92859838e-02   8.82286911e-02   1.88411109e-01
   2.11918665e-01   1.35100293e-01   2.71404031e-01  -2.04929852e-02
   2.78162748e-01   2.55930791e-01   5.66850964e-02  -4.92616346e-02
   4.20378517e-02   3.67132244e-02   1.18991678e-01   2.08621664e-01
   8.30118096e-02   4.63128592e-02   1.72467381e-01   2.08386960e-01
   2.09317185e-02  -9.14845716e-02   2.55859830e-01  -1.16752169e-02
   2.32545923e-01   2.20468954e-01   1.37154610e-01   6.58246925e-02
   2.00251037e-01   1.81200474e-03   5.39732331e-02   1.34733282e-01
  -4.21934755e-03   3.57510840e-01   3.49796758e-02   3.18340727e-01
   1.53523436e-01  -6.84982568e-02   1.18521865e-01   1.46500547e-01
   3.44573408e-01   1.34395099e-01   1.19096954e-01   2.32965339e-01
   3.93928233e-01   1.50820310e-01   1.97450380e-01   6.06068240e-02
   6.90928326e-02   1.69967649e-02   2.80257948e-01   1.28551206e-01
  -7.65747560e-02   5.24285636e-02   4.26535845e-01   2.41638949e-01
   1.93231751e-01   1.47666759e-01   3.12781721e-01   2.80117983e-01
   2.29734944e-01   1.18325263e-01   1.72010854e-01   3.86053458e-02
   2.73964412e-01   2.91562778e-01   2.16578587e-02   7.93291253e-02
   1.83240324e-01]
MSE:  4.4330465939
```

1.2a) Normalization does affect Beta values and MSE. The most notable change occurs in Batch Gradient with normalization/ without normalization. The MSE without normalization is 5.50765097196, while the MSE with normalization is a staggering 440.992321592. This was also accompanied by larger values of beta. This could be due to the fact that normalization over compensates for the different ranges of feature values. For example sometimes you don't want to use normalization if the units of your measurement are meaningful. The other MSE errors values do not fluctuate much with/without normalization which could be indicative of the fact data has similar value ranges.

1.2b) Similarly, Z-score normalization had variable effects on the data, ranging from significant to insignificant. This is likely due to the distributions observed during the different training methods as the z-score enables us to compare two scores that are from different normal distributions.

2a) $$L = \left(0 - \log\left(1+e^{\left(\begin{bmatrix}1\\60\\155\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix}\right)}\right)\right) + \left(1\cdot\begin{bmatrix}1\\64\\135\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix} - \log\left(1+e^{\left(\begin{bmatrix}1\\64\\135\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix}\right)}\right)\right)$$

$$+ \left(1\cdot\begin{bmatrix}1\\73\\170\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix} - \log\left(1+e^{\left(\begin{bmatrix}1\\73\\170\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix}\right)}\right)\right)$$

$$\boxed{\begin{aligned} L = &-\log\left(1+e^{(\beta_0+60\beta_1+155\beta_2)}\right) + \left(\beta_0+64\beta_1+135\beta_2 - \log\left(1+e^{(\beta_0+64\beta_1+135\beta_2)}\right)\right) \\ &+ \left(\beta_0+73\beta_1+170\beta_2 - \log\left(1+e^{(\beta_0+73\beta_1+170\beta_2)}\right)\right)\end{aligned}}$$

2b. $$\frac{\partial L(\beta)}{\beta_0} = 1(0-p(x_1;\beta)) + 1(1-p(x_2;\beta)) + 1(1-p(x_3;\beta))$$

$$X = \begin{bmatrix}1 & 60 & 155\\ 1 & 64 & 135\\ 1 & 73 & 170\end{bmatrix}$$

$$\frac{\partial L(\beta)}{\beta_1} = 60(0-p(x_1;\beta)) + 64(1-p(x_2;\beta)) + 73(1-p(x_3;\beta))$$

$$Y = \begin{bmatrix}0\\1\\1\end{bmatrix}$$

$$\frac{\partial L(\beta)}{\beta_2} = 155(0-p(x_1;\beta)) + 135(1-p(x_2;\beta)) + 170(1-p(x_i;\beta))$$

where $$p(x_i;\beta) = \frac{e^{\beta^T x_i}}{1+e^{\beta^T x_i}}$$

$$x_1 = \begin{bmatrix}1\\1\\1\end{bmatrix},\quad x_2 = \begin{bmatrix}60\\64\\73\end{bmatrix}\quad x_3 = \begin{bmatrix}155\\135\\170\end{bmatrix}$$

2c) Hessian:

$$\frac{\partial L(\beta)}{\partial \beta_{1j}\,\partial \beta_{1n}} = -\sum_{i=1}^{N} x_{ij} x_{in}\, p(x_i;\beta)(1-p(x_i;\beta))$$

$$\frac{\partial L(\beta)}{\partial \beta_{0j}\,\partial \beta_{1n}} = 1\cdot 1\, p(x_i;\beta)(1-p(x_i;\beta)) + 1\cdot 60\, p(x_i;\beta)(1-p(x_i;\beta)) + 1\cdot 155(x_i;\beta)(1-p(x_i;\beta)) + 60\cdot 1\, p(x_i;\beta)(1-p(x_i;\beta)) + 60\cdot 60\, p(x_i;\beta)(1-p(x_i;\beta)) + 60\cdot 155\, p(x_i;\beta)(1-p(x_i;\beta)) + 155\cdot 1\, p(x_i;\beta)(1-p(x_i;\beta)) + 155\cdot 60\, p(x_i;\beta)(1-p(x_i;\beta)) \; 155\cdot 155\, p(x_i;\beta)(1-p(x_i;\beta))$$

Let $f(x,\beta) = p(x_i;\beta)(1-p(x_i;\beta))$

$$\frac{\partial L(\beta)}{\partial \beta_{1j}\,\partial \beta_{1n}} = 1\cdot 1\, f(x,\beta) + 1\cdot 64\, f(x,\beta) + 1\cdot 135\, f(x,\beta) + 64\cdot 1\, f(x,\beta) + 64\cdot 64\, f(x,\beta) + 64\cdot 135\, f(x,\beta) + 135\cdot 1\, f(x,\beta) + 135\cdot 64\, f(x,\beta) + 135\cdot 135\, f(x,\beta)$$

$$\frac{\partial L(\beta)}{\partial \beta_{2j}\,\partial \beta_{2n}} = 1\cdot 1\, f(x,\beta) + 1\cdot 73\, f(x,\beta) + 1\cdot 170\, f(x,\beta) + 73\cdot 1\, f(x,\beta) + 73\cdot 73\, f(x,\beta) + 73\cdot 170\, f(x,\beta) + 170\cdot 1\, f(x,\beta) + 170\cdot 73\, f(x,\beta) + 170\cdot 170\, f(x,\beta)$$

The circled portions of the above equations are the entries of the 3x3 Hessian matrix

3.1 - "handicapped infants = yes": 6 Dems, 2 Rep

info([6, 2]) = entropy($\frac{6}{8}$, $\frac{2}{8}$)

$= -\frac{6}{8}\log\left(\frac{6}{8}\right) - \frac{2}{8}\log\left(\frac{2}{8}\right) = .811 \rightarrow .811\left(\frac{8}{20}\right) + .918\left(\frac{12}{20}\right)$

$= .8752$

- "handicapped infants = no": 4 Dems, 8 Rep

info([4, 8]) = entropy($\frac{4}{12}$, $\frac{8}{12}$) $= -\frac{4}{12}\log\left(\frac{4}{12}\right) - \frac{8}{12}\log\left(\frac{8}{12}\right) = .918$

- "water project = yes": 4 Dems, 6 Rep

info([4, 6]) = entropy = ($\frac{4}{10}$, $\frac{6}{10}$) $= -\frac{4}{10}\log\left(\frac{4}{10}\right) - \frac{6}{10}\log\left(\frac{6}{10}\right) = .97$

$\rightarrow .97\left(\frac{10}{20}\right) + .97\left(\frac{10}{20}\right) = .97$

- "water project = no": 6 Dems, 4 Rep

info([6, 4]) = entropy($\frac{6}{10}$, $\frac{4}{10}$) $= -\frac{6}{10}\log\left(\frac{6}{10}\right) - \frac{4}{10}\log\left(\frac{4}{10}\right)$

$= .97$

- "budget resolution = yes": 9 Dems, 2 Rep

info([9, 2]) = entropy($\frac{9}{11}$, $\frac{2}{11}$)

$= -\frac{9}{11}\log\left(\frac{9}{11}\right) - \frac{2}{11}\log\left(\frac{2}{11}\right) = .684$

$\rightarrow .684\left(\frac{11}{20}\right) + .50\left(\frac{9}{20}\right)$

$= .60$

- "budget resolution = no": 1 Dem, 8 Rep

info([1, 8]) = entropy($\frac{1}{9}$, $\frac{8}{9}$)

$= -\frac{1}{9}\log\left(\frac{1}{9}\right) - \frac{8}{9}\log\left(\frac{8}{9}\right)$

$= .50$

gain("Handicapped Infants") = info([10,10]) - info([6,2],[4,8])

$= -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) - .8752$

$= 1 - .8752 = \boxed{.1248}$

gain("water project") = 1 - info([4,6],[6,4])

$= 1 - .97 = \boxed{.03}$

gain("budget resolution") = 1 - info([9,2],[1,8])

$= 1 - .60 = \boxed{.4}$

$0.4 > 0.1248 > 0.03$

So split on "budget resolution"

budget resolution = yes → 9 Dems, 2 Rep (11 total)

"Handicapped Infants = Yes": 6 Dems, 1 Rep

info([6,1]) = entropy(6/7, 1/7) $= -\frac{6}{7}\log\left(\frac{6}{7}\right) - \frac{1}{7}\log\left(\frac{1}{7}\right)$

$= .59$

"Handicapped Infants = No": 3 Dems, 1 Rep

info([3,1]) = entropy(3/4, 1/4) $= -\frac{3}{4}\log\left(\frac{3}{4}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right)$

$= .811$

$\rightarrow .59\left(\frac{7}{11}\right) + .811\left(\frac{4}{11}\right) = .67$ ← expected info

"Water Project = Yes": → same as handicapped infants = no

→ .811

"Water Project = No": → same as handicapped infants = yes

→ .59

expected info → .67

Branch on either one

"budget resolution = no" → 1 Dem, 8 Reps (9 total)

"Handicapped Infants = Yes": 0 Dems, 1 Rep

info([0,1]) = entropy(0/1, 1/1) = 0 − 1 log(1) = $\boxed{0}$

"Handicapped Infants = No": 1 Dem, 7 Rep

info([1,7]) = entropy(1/8, 7/8) = $-\frac{1}{8}\log\left(\frac{1}{8}\right) - \frac{7}{8}\log\left(\frac{7}{8}\right)$

Info([1,8]) − info([0,1],[1,7]) = .503 − .48 = $\boxed{.023}$ = .54

expected info → $0\left(\frac{1}{9}\right) + .54\left(\frac{8}{9}\right) = \boxed{.48}$

"Water Project = Yes": 1 Dems, 5 Rep

info([1,5]) = entropy(1/6, 5/6) = $-\frac{1}{6}\log\left(\frac{1}{6}\right) - \frac{5}{6}\log\left(\frac{5}{6}\right)$

= .65

"Water Project = No": 0 Dem, 3 Rep

info([0,3]) = entropy(0/3, 3/3) = 0 − 1 log(1) = 0

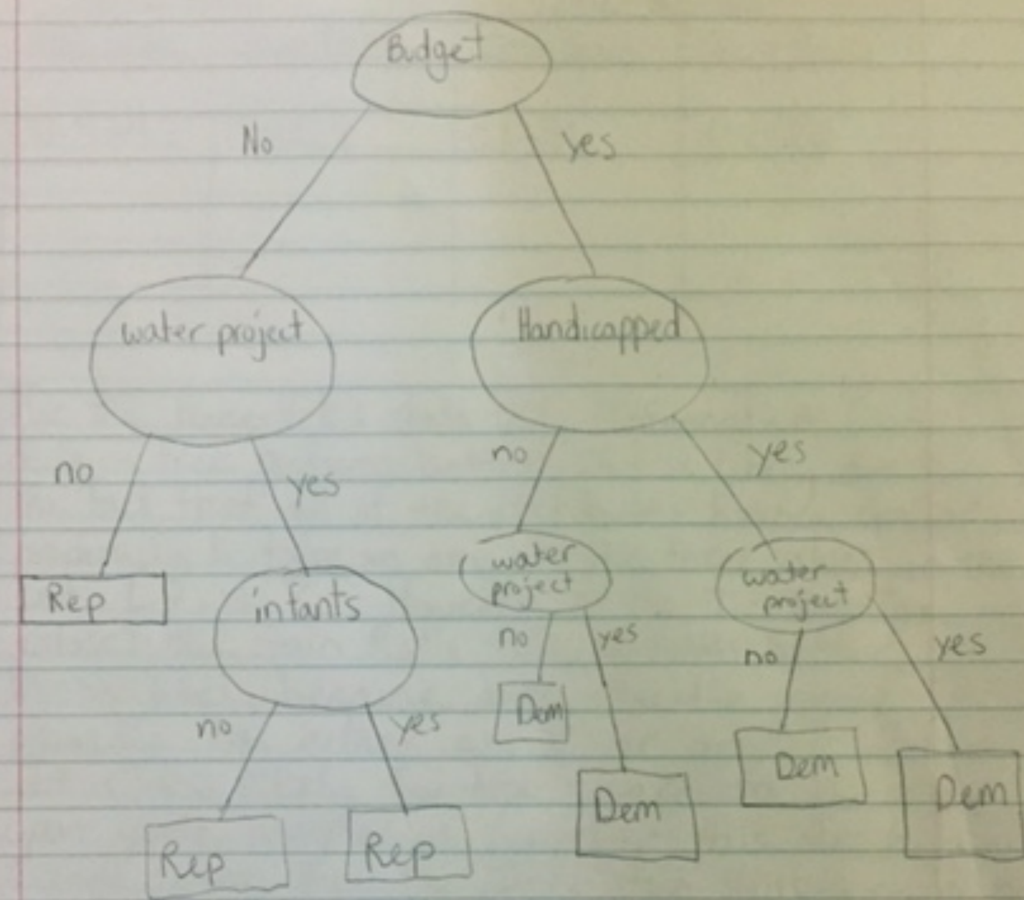
expected info → $.65\left(\frac{6}{9}\right) + 0 = .43$

Info [1,8] − info([1,5],[0,3]) =

entropy(1/9, 8/9) − .43

$-\frac{1}{9}\log\left(\frac{1}{9}\right) - \frac{8}{9}\log\left(\frac{8}{9}\right) - .43 = \boxed{0.073}$

Branch on "Water Project"



Budget
No
yes
water project
Handicapped
no
yes
no
yes
Rep
infants
water project
water project
no
yes
no
yes
no
yes
Dem
Dem
Dem
Dem
Rep
Rep

	House-Votes		Tic-Tac-Toe	
3.2b)	Info Gain	Gain Ratio	Info Gain	Gain Ratio
5-CV	0.9241	0.9218	0.8371	0.8392

For the House-Votes data set, Information Gain decision tree performs better. This is likely due to the fact that all of the attributes have a similar propensity to take on any of the three values in the set [y, n, ?]. However, on the Tic-Tac-Toe dataset the Gain Ratio tree performs better. This is likely because a tic-tac-toe square usually has either an "x" or an "o" in it. It is less likely for the square to be empty upon game completion, however this does happen in some cases. Thus, the Info Gain measure could be biased to attributes that often take on any of the three values from the set ["x", "o", "b"].