1.1

a. $y = \beta_0 + x_1\beta_1 = \mathbf{X^T\beta}$

Closed-form: $\boldsymbol{\beta}=(X^TX)^{-1}X^TY$

$$X = \begin{bmatrix} 1 & 60 \\ 1 & 70 \\ 1 & 62 \\ 1 & 72 \\ 1 & 65 \end{bmatrix} \quad Y = \begin{bmatrix} 130 \\ 155 \\ 125 \\ 162 \\ 150 \end{bmatrix}$$

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

$$\text{determinant}(X^TX) = 1/(5*21753 - 329*329) = 1/524$$

$$(X^TX)^{-1} = 1/524 * \begin{bmatrix} 21753 & -329 \\ -329 & 5 \end{bmatrix}$$

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

$$\boldsymbol{\beta}=(X^TX)^{-1}X^TY = 1/524 * \begin{bmatrix} 21753 & -329 \\ -329 & 5 \end{bmatrix} * \begin{bmatrix} 722 \\ 47814 \end{bmatrix} = 1/524 * \begin{bmatrix} -25140 \\ 1532 \end{bmatrix} = \begin{bmatrix} -47.977 \\ 2.924 \end{bmatrix}$$

b. $h(x) = \mathbf{X\beta}$

$$= \begin{bmatrix} 1 & 60 \\ 1 & 70 \\ 1 & 62 \\ 1 & 72 \\ 1 & 65 \end{bmatrix} * \begin{bmatrix} -47.977 \\ 2.924 \end{bmatrix}$$

$$= \begin{bmatrix} 127.463 \\ 156.703 \\ 133.311 \\ 163.991 \\ 142.083 \end{bmatrix}$$

1.2

a.

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:
[-4.26544322 -1.05015978  4.04012051  2.44527069  2.63191327  1.94060747
 1.728485    1.36856652  0.95094023  0.68389666  0.75374135  0.78885674
 0.32751618  0.83431982  1.12581097  0.35960437  0.15084528  0.82861544
 0.68665784  0.6475998   0.57229887  0.4264666   0.49805077  0.89400043
 0.62321048  0.73016534  0.56915696  0.60028539  0.43842633  0.48429968
 0.0784136   0.95259901  0.49735484  0.41503377  0.93140872  0.97478587
 0.33326308  0.95249924  0.8223191   0.44991619  0.34423522  0.12169874
 0.75292086  0.05869686  0.15551465  0.34062733  0.57956898  0.43709462
 0.4243603   0.094368    0.82545051  0.12199582  0.5110402   0.68768422
 0.38358163  0.64464014  0.70144089  0.96997327  0.17916647  0.09360594
 0.52986936  0.37624458  0.52565223  0.83760645  0.96211074  0.48751837
 0.02853175  0.56336924  0.58087998  0.52746266  0.79763062  0.96907893
 0.60342546  0.04539628  0.11945264  0.59142365  0.30821628  0.26502191
 0.68563508  0.36115615  0.45456424  0.27660829  0.75482827  0.07795255
 0.47900789  0.70839575  0.05625217  0.99532464  0.30570233  0.42464403
 0.81619259  0.8668387   0.3882989   0.34275747  0.33937469  0.74124369
 0.85797067  0.45447504  0.59352046  0.49668153  0.74885129]
MSE:  19.0042206311
-----------------------------------------------
Stochastic Gradient Without Normalization
Beta:
[ 0.08657766  0.09592712  0.27397496  0.47739931  0.83639798  0.78203068
 0.11930966  0.3580488   0.41381203  0.67981587  0.34226278  0.60186583
```

0.23194178 0.66322631 0.42978908 0.87779128 0.09920388 0.32555479
0.57356113 0.75368124 0.19965248 0.86597887 0.54826947 0.55837936
0.62265264 0.1875552 0.41269964 0.40106274 0.09307619 0.29259798
0.53652515 0.15883613 0.6176283 0.16033609 0.78131556 0.87408055
0.62483647 0.13501113 0.8566379 0.00513402 0.53858349 0.62474698
0.16580628 0.60945349 0.13322909 0.18299517 0.12345151 0.5290512
0.70869192 0.50190491 0.04995951 0.21391733 0.32433041 0.71380968
0.49178307 0.32646567 0.17296836 0.27458549 0.03746842 0.17627834
0.96355351 0.53513032 0.63813549 0.33048459 0.13376107 0.66083466
0.77843902 0.98320116 0.2745575 0.06350764 0.16297118 0.47699384
0.17616956 0.72139132 0.15302852 0.3810725 0.27214754 0.63470034
0.74711616 0.47121784 0.47102078 0.43700858 0.40909383 0.76184832
0.61479098 0.50667356 0.17179854 0.82312338 0.85614079 0.13254583
0.91949029 0.63138745 0.63251464 0.76056908 0.23560187 0.31305743
0.94749162 0.36654411 0.60002645 0.05523669 0.23583955]
MSE: 5.63248422192

They are not same because the result really depends on the initial point. So, the result may be a local minima rather than the absolute minima. In addition, the learning rate alpha also plays a big role in the batch and stochastic gradient. If alpha is too small, the calculation is accurate but it moves really slowly. In contrast, if alpha is too big, the result may not get to the minima.

b.
Closed Form With Normalization
Beta:
[2.27729720e+01 1.53267685e-01 1.85400036e-01 1.20001101e-01
-5.02894960e-03 8.91855522e-02 2.85477509e-01 1.40249729e-01
7.58001703e-02 1.29653087e-02 9.40114997e-02 3.31501951e-01
8.48405150e-02 1.19998020e-01 2.42101087e-01 -1.70904428e-02
1.37119556e-01 1.35350218e-01 1.41619004e-03 5.96423043e-03
1.15830867e-01 2.84837752e-01 4.40248244e-02 2.49185633e-01
1.20285952e-01 -1.97966211e-02 9.78939759e-02 8.05403060e-02
-1.21241111e-02 -6.77059821e-02 4.42940642e-02 1.07814670e-01
2.27982170e-01 4.72154203e-02 7.98729034e-02 1.82957097e-01
2.10609705e-01 1.62079663e-01 2.74455584e-01 -1.24456123e-02
2.32346197e-01 2.68821067e-01 3.49745502e-02 -5.73174263e-02
2.74558199e-02 3.22366923e-02 1.03219840e-01 2.23792899e-01
1.12445398e-01 3.34223468e-02 1.96611852e-01 2.04171370e-01
1.61259528e-02 -7.12316220e-02 2.51757075e-01 -3.88735810e-03
2.31055679e-01 2.65481860e-01 1.14239087e-01 7.19519080e-02
2.03225977e-01 -2.77922653e-03 5.10043840e-02 1.31478537e-01
7.74623329e-03 3.36781203e-01 1.19518825e-02 2.98145298e-01
1.64253970e-01 -8.57326109e-02 9.04810592e-02 1.57878654e-01

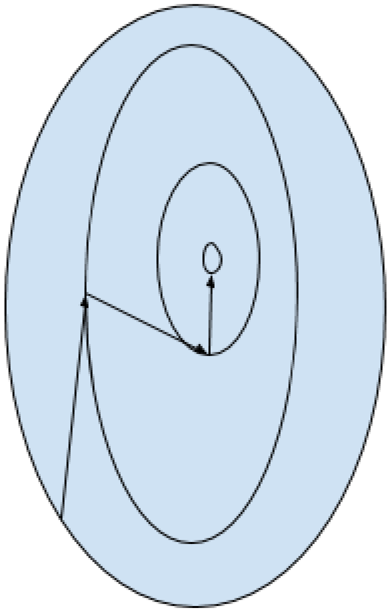
3.30578812e-01 1.58142457e-01 1.17519641e-01 2.66603450e-01
3.90619100e-01 1.54573813e-01 1.82230684e-01 6.25165215e-02
6.11873098e-02 -1.74345404e-03 2.34361003e-01 1.35158424e-01
-7.34879378e-02 5.72871764e-02 4.02966409e-01 2.50642329e-01
1.87572968e-01 1.57855445e-01 3.18225717e-01 2.66144412e-01
2.29911686e-01 8.53225833e-02 1.56706806e-01 6.57087894e-02
2.52781623e-01 2.90068806e-01 2.33284776e-02 9.01229905e-02
2.03998476e-01]
MSE: 4.40454594906

Batch Gradient With Normalization
Beta:
[2.27729720e+01 1.18270874e-01 -6.70689874e-02 5.84364047e-02
-1.43693497e-01 2.67318941e-01 2.55238247e-01 2.66945142e-01
-1.00626911e-01 3.06082147e-01 2.91216224e-01 5.54988083e-01
1.26349480e-01 2.23569559e-01 1.18696212e-01 -9.38152218e-02
2.00319392e-01 1.93275179e-01 8.34456780e-02 -1.05826075e-01
2.77902840e-01 1.18145664e-01 2.25983736e-01 1.80318246e-01
5.55147896e-02 -2.05065532e-01 1.60028822e-02 5.13018534e-02
-4.01900991e-03 3.83745988e-02 -7.44546755e-02 1.39294197e-01
1.86395871e-01 -2.09813758e-01 6.14532770e-02 2.61934407e-01
2.01147256e-01 9.81781347e-03 5.03858371e-01 -1.55629409e-01
2.89877209e-01 3.80479123e-01 -8.17826678e-02 -1.05370663e-01
2.66897324e-01 -1.45159381e-01 -2.92238798e-02 1.61526393e-01
9.42719409e-02 -4.12829754e-03 1.32292093e-02 2.18075965e-01
-1.60507111e-02 -6.61264036e-02 8.53672092e-02 -4.44122534e-03
2.44099793e-01 4.67458750e-01 8.15690385e-02 -2.17259080e-02
1.93024739e-01 1.42943077e-01 -9.32003116e-02 7.37113248e-02
2.50394314e-01 5.32532617e-01 3.34242380e-03 1.19547733e-01
8.80152963e-02 -3.77620167e-02 1.13974650e-01 1.66365091e-01
4.22770534e-01 2.37985275e-01 8.95117257e-02 2.35282104e-01
4.02622475e-01 7.57736028e-02 1.32994092e-01 1.60505471e-01
-1.62275192e-02 1.23759506e-02 1.87420153e-01 2.03629831e-01
8.98470967e-02 6.91245387e-02 5.53894457e-01 2.58369698e-01
1.72906638e-01 1.07619317e-01 5.05184238e-01 2.87324921e-01
3.79842237e-01 1.75791497e-01 1.21601176e-01 1.26413483e-01
2.61368731e-01 4.33213709e-01 1.29782725e-01 9.23276488e-02
2.63774112e-01]
MSE: 5.45767269472

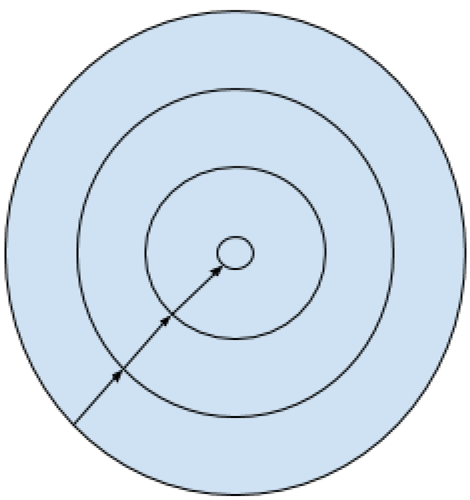
Stochastic Gradient With Normalization
Beta:
[1.51130338e+01 8.96761898e-02 2.72731778e-01 3.41131537e-01

```
   8.14136161e-02   2.77806269e-01   3.20483294e-01   4.88012832e-01
   2.40131943e-01   1.26617135e-01   4.29576512e-01   7.06590129e-01
   2.85446544e-01   4.02097388e-01   4.15573062e-01   2.93776568e-01
   2.89404056e-01   3.17064331e-01   1.41024398e-01   2.75900036e-02
   1.91734231e-01   3.36698021e-01   1.37000889e-01   3.77753872e-01
   2.75616496e-01   3.64343296e-01   7.02329757e-02   3.27011102e-01
   3.40366440e-01   2.14520508e-01   1.90477747e-01   9.88705051e-02
   5.02399308e-01   2.68547693e-01   1.19222390e-01   6.06151711e-01
   1.06878978e-01   2.79539940e-01   4.46556749e-01   2.51067839e-01
   3.21233342e-01   4.72949045e-01   2.70639625e-01   3.20295165e-01
   3.85027187e-01   4.11476984e-01   3.10814296e-01   3.90620879e-01
   2.31832959e-01   7.98103264e-02   8.95299157e-02   1.55990154e-01
   1.94289215e-01   9.96308981e-03   5.76078445e-01   2.42972240e-01
   1.67346517e-01   8.92946427e-02   1.93719860e-01   3.37219834e-01
   1.69865324e-01   1.59203039e-01  -1.36275761e-02   1.91725579e-01
   2.06952657e-01   4.09534126e-01   2.32059552e-01   3.91859273e-01
   3.95361719e-01  -1.15329187e-01   2.64941302e-01   1.68852610e-01
   2.64247892e-01   3.19216242e-01   1.67697622e-01   3.96966248e-01
   3.71184046e-01   1.24896782e-01   4.06672916e-01   2.54679917e-01
   2.40735443e-01  -4.18811004e-02   4.44981530e-01   2.36699948e-01
  -1.10167243e-02   4.49995529e-01   5.48067470e-01   2.67754775e-01
   4.25609972e-01   4.64348004e-01   3.97615652e-01   5.40729593e-01
   3.11333222e-01   3.77947393e-01   2.51258387e-01   1.06577059e-01
   5.38377267e-01   1.01539425e-01   2.94195249e-02   5.78277500e-03
   2.38455459e-01]
MSE:  64.5916473389
-----------------------------------------------
Normalization affects both beta and MSE.
```

Because before the data is normalized, if the range of one attribute is way bigger than the range of another one, then the contours of the graph would be oval, like:



After normalization, the coefficients are close. So, the contours of would be close to circle:



So, the path is much smoother than the one before normalization and it is easier to get the optimal solution.

2.

a. $L = \Sigma_i(y_i x_i^T\beta - \log(1+\exp\{x_i^T\beta\}))$

$= 0 - \log(1+ \exp\{\beta_0 + 60\beta_1 + 155\beta_2\}) + \beta_0 + 64\beta_1 + 135\beta_2 - \log(1+ \exp\{\beta_0 + 64\beta_1 + 135\beta_2\}) + \beta_0 + 73\beta_1 + 170\beta_2 - \log(1+ \exp\{\beta_0 + 73\beta_1 + 170\beta_2\})$

$=2\beta_0 + 137\beta_1 + 305\beta_2 - \log(1+ \exp\{\beta_0 + 60\beta_1 + 155\beta_2\}) - \log(1+ \exp\{\beta_0 + 64\beta_1 + 135\beta_2\}) - \log(1+ \exp\{\beta_0 + 73\beta_1 + 170\beta_2\})$

b. $\nabla L(\beta_0) = 2 - 1 + \frac{1}{1+ exp\{\beta0 + 60\beta1 + 155\beta2\}} - 1 + \frac{1}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}} - 1 + \frac{1}{1+ exp\{\beta0 + 73\beta1 + 170\beta2\}}$

$= -1 + \frac{1}{1+ exp\{\beta0 + 60\beta1 + 155\beta2\}} + \frac{1}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}} + \frac{1}{1+ exp\{\beta0 + 73\beta1 + 170\beta2\}}$

$\nabla L(\beta_1) = 137 - 60 + \frac{60}{1+ exp\{\beta0 + 60\beta1 + 155\beta2\}} - 64 + \frac{64}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}} - 73 + \frac{73}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}}$

$= -60 + \frac{60}{1+ exp\{\beta0 + 60\beta1 + 155\beta2\}} + \frac{64}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}} + \frac{73}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}}$

$\nabla L(\beta_2) = 305 - 155 + \frac{155}{1+ exp\{\beta0 + 60\beta1 + 155\beta2\}} - 135 + \frac{135}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}} - 170 + \frac{170}{1+ exp\{\beta0 + 73\beta1 + 170\beta2\}}$

$= -155 + \frac{155}{1+ exp\{\beta0 + 60\beta1 + 155\beta2\}} + \frac{135}{1+ exp\{\beta0 + 64\beta1 + 135\beta2\}} + \frac{170}{1+ exp\{\beta0 + 73\beta1 + 170\beta2\}}$

c. $H_{\beta0}= [- \frac{1}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{1}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} - \frac{1}{2+exp\{\beta0 + 73\beta1 + 170\beta2\} +exp\{-\beta0 - 73\beta1 - 170\beta2\}},$

$- \frac{60}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{64}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} - \frac{73}{2+exp\{\beta0 + 73\beta1 + 170\beta2\} +exp\{-\beta0 - 73\beta1 - 170\beta2\}},$

$- \frac{155}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{135}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} - \frac{170}{2+exp\{\beta0 + 73\beta1 + 170\beta2\} +exp\{-\beta0 - 73\beta1 - 170\beta2\}}]$

$H_{\beta1}= [- \frac{60}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{64}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} - \frac{73}{2+exp\{\beta0 + 73\beta1 + 170\beta2\} +exp\{-\beta0 - 73\beta1 - 170\beta2\}},$

$- \frac{3600}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{4096}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} - \frac{5329}{2+exp\{\beta0 + 73\beta1 + 170\beta2\} +exp\{-\beta0 - 73\beta1 - 170\beta2\}},$

$- \frac{9300}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{8460}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} - \frac{12410}{2+exp\{\beta0 + 73\beta1 + 170\beta2\} +exp\{-\beta0 - 73\beta1 - 170\beta2\}}]$

$H_{\beta2} = [- \frac{155}{2+exp\{\beta0 + 60\beta1 + 155\beta2\} +exp\{-\beta0 - 60\beta1 - 155\beta2\}} - \frac{135}{2+exp\{\beta0 + 64\beta1 + 135\beta2\} +exp\{-\beta0 - 64\beta1 - 135\beta2\}} -$

$\frac{170}{2+exp\{\beta 0 + 73\beta 1 + 170\beta 2\} + exp\{-\beta 0 - 73\beta 1 - 170\beta 2\}}$,

$- \frac{9300}{2+exp\{\beta 0 + 60\beta 1 + 155\beta 2\} + exp\{-\beta 0 - 60\beta 1 - 155\beta 2\}} - \frac{8640}{2+exp\{\beta 0 + 64\beta 1 + 135\beta 2\} + exp\{-\beta 0 - 64\beta 1 - 135\beta 2\}} - \frac{12410}{2+exp\{\beta 0 + 73\beta 1 + 170\beta 2\} + exp\{-\beta 0 - 73\beta 1 - 170\beta 2\}}$,

$- \frac{24025}{2+exp\{\beta 0 + 60\beta 1 + 155\beta 2\} + exp\{-\beta 0 - 60\beta 1 - 155\beta 2\}} - \frac{18225}{2+exp\{\beta 0 + 64\beta 1 + 135\beta 2\} + exp\{-\beta 0 - 64\beta 1 - 135\beta 2\}} - \frac{28900}{2+exp\{\beta 0 + 73\beta 1 + 170\beta 2\} + exp\{-\beta 0 - 73\beta 1 - 170\beta 2\}}$]

3. 1

h = Vote for handicapped-infants
w = Vote for water-project-cost-sharing
b = Vote for budget-resolution-adoption

Step 1:

info([10,10]) = 1

h:
Yes: [6, 2]
No: [4, 8]
info([6, 2]) = 0.811
info([4, 8]) = 0.918
info([6,2], [4,8]) = ⅖*0.811 + ⅗ * 0.918 = 0.8752
gain(h) = info([10, 10]) - info([6,2], [4,8]) = 1 - 0.8752 = 0.1248

w:
Yes: [6, 4]
No: [6, 4]
info([6,4]) = 0.971
info([6,4], [6,4]) = 0.971
gain(w) = 1 - 0.971 = 0.029

b:
Yes: [9, 2]
No: [8, 1]
info([9, 2]) = 0.684
info([8, 1]) = 0.503
info([9, 2], [8, 1]) = 11/20 * 0.684 + 9/20 * 0.503 = 0.603
gain(b) = 1 - 0.603 = 0.397
So, b is a better choice

Step 2:

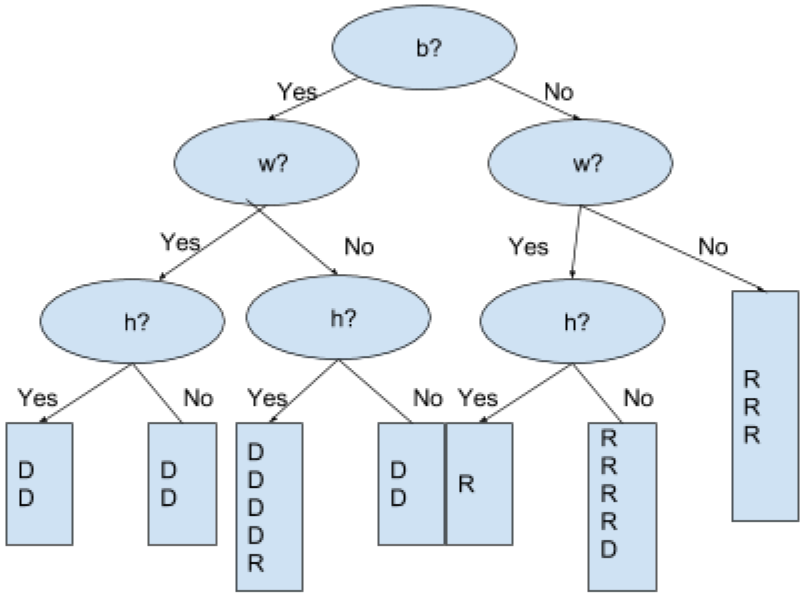
When b=Yes
info([9, 2]) = 0.684
h:
Yes: [6, 1]
No: [3, 1]

info([6, 1]) = 0.592
info([3, 1]) = 0.811
info([6,1],[3,1]) = 7/11*0.592 + 4/11*0.811 = 0.672
gain(h) = 0.684 - 0.672 = 0.012

w:
Yes: [3, 1]
No: [6, 1]
gain(w) = 0.012
When b=Yes, h and w are same

When b = No
info([8,1]) = 0.503
h:
Yes:[1, 0]
No: [7, 1]
info([1,0]) = 0
info([7,1]) = 0.544
info([1,0],[7,1]) = 1/9 * 0 + 8/9 * 0.544 = 0.484
gain(h) = 0.503 - 0.484 = 0.019

w:
Yes: [5, 1]
No:[3,0]
info([5,1]) = 0.650
info([3,0]) = 0
info([5,1], [3,0]) = 0.650 * ⅔ + 0 = 0.433
gain(w) = 0.503 - 0.433 = 0.070
When b = No, w is better



So, the tree is: