Homework 6 Solutions

# Exercise 1

1. f[(3,2)] =

|  |  |  |  |
| --- | --- | --- | --- |
| 1-NN | 2-NN | 3-NN | 4-NN |
| 30 | 25 | 15 | 12 |

* 1. 1-NN. Since *p* is nearest to the point of maximum value, and all values are distinct, 2-NN can only decrease the estimate.
  2. Unable to determine. The third point could cause an increase or decrease from the 2-NN estimate, depending on whether its value is greater or lesser than the second point.
  3. 1-NN. The maximum will be strictly greater than the average, since all points are distinct (assuming n greater than 1).
  4. Unable to determine. The average of the first two points could be significantly lower or higher than the true average.

1. Precisely the set of points equidistant to both (0,0) and (2,0), which is the vertical line through (1,0).
2. Non-weighted 2-NN predicts 11.5 for each point. Weighted 2-NN predicts 3 for (0,0) and 20 for (2,0), the same as their values in the table.

# Exercise 2

|  |  |  |
| --- | --- | --- |
| Mary \ John | Advocate A | Advocate B |
| Advocate A | M = +2, J = +2 | M = +2, J = -6 |
| Advocate B | M = -6, J = +2 | M = +5, J=+5 |

* 1. No. To be a prisoner’s dilemma, each player must have an incentive to defect *regardless* of the other player’s choice. In this case, advocating A is like defecting, but if Mary advocates B, then John also has an incentive to advocate B. Thus there are 2 Nash equilibria in this game: the top left and the bottom right. If this were a prisoner’s dilemma, it would only be the top left. To make this a true prisoner’s dilemma, the reward for advocating A when the other player advocates B must be higher than +5.
  2. F1 = X, F2 = Y is the only nash equilibrium. This means that if Firm 1 releases product X, then Firm 2 has the strongest incentive to release product Y. Similarly, if Firm 2 releases product Y, then Firm 1 has the strongest incentive to release product X.
  3. It is not Pareto optimal. F1 = Z, F2 = Z is Pareto optimal, because from that square, neither firm can become better off without the other becoming worse off. F1 = Y, F2 = Y and F1 = Z, F2 = Y are also Pareto optimal.

# Exercise 3

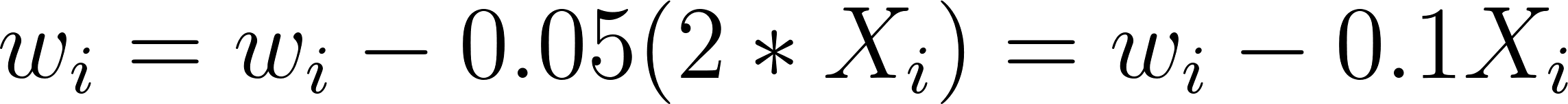
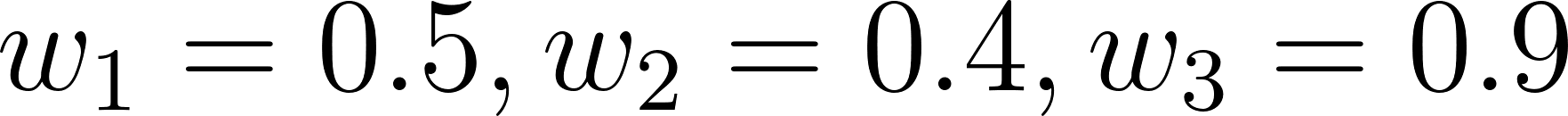


If you have information, you expect 0.8\*($-1000)+0.2\*($-5000) = $ -1800 in maintenance. If you don’t have information, you expect 0.5\*($-1000)+0.5\*($-5000) = $-3000 in maintenance. Thus, the information is worth $1200 in expectation.

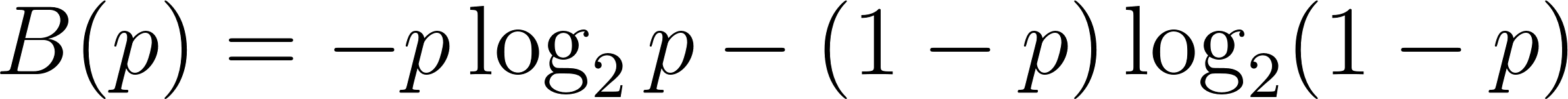
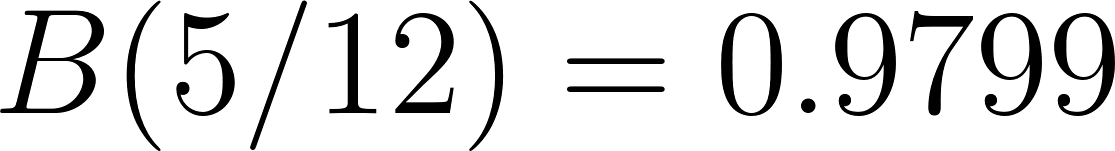
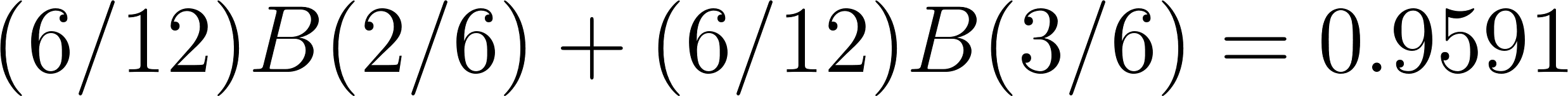
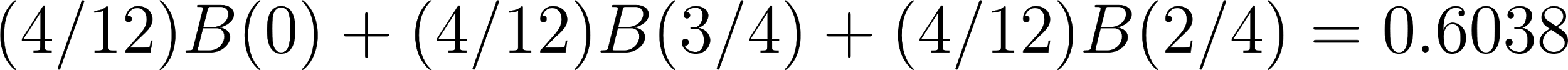
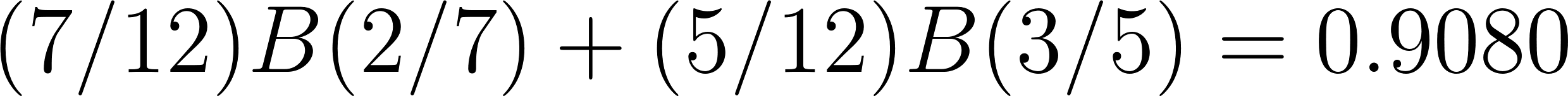
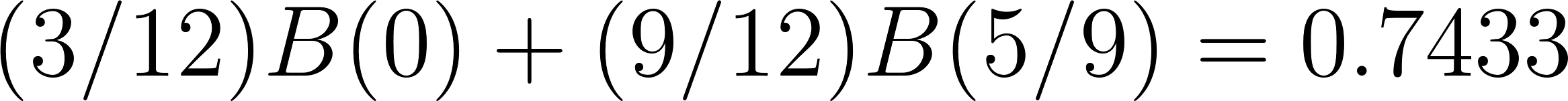
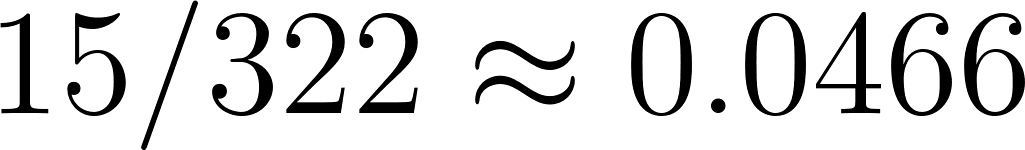
# Exercise 4

We can obtain the gradient of the loss with respect to each weight as follows:



With learning rate 0.05, backpropagation will update with the formula . Thus after backpropagation we have .

# Exercise 5

1. 2\*3\*2\*2 = 24. (Multiply the number of categories for each environmental feature).
2. Following the notation in the book, let  be the entropy function. The initial entropy is .  
   Splitting on “Slope”:   
   Splitting on “Terrain”: .  
   Splitting on “Temperature”: .  
   Splitting on “Humidity”: .  
     
   Information gain: Slope = 0.0208, Terrain = 0.3761, Temperature = 0.07190, Humidity = 0.2366
3. 
4.   
     
   This tree is not discovered by the max information gain algorithm because that algorithm is greedy and non-optimal. Temperature appears to have a low information gain because each side of the decision tree with temperature at the root is fairly balanced in positive and negative examples in the given data.
5. If the tree misclassifies 4 environments in total, the chance that it would still perfectly classify a random set of 12 distinct points is 20 choose 12 divided by 24 choose 12 (the number of size-12 subsets that would be perfectly classified, divided by the total number of size-12 subsets). This is .

# Exercise 6

1. a: a, c  
   b: b, c  
   c: a, b, c, d  
   d: c, d
2. We need to calculate P(end state | no wall, attempt to go north). By the definition of conditional probability, this is equal to P(end state, no wall | attempt to go north) / P(no wall | attempt to go north). We can calculate P(end state, no wall | attempt to go north) = sum over start state of P(end state, no wall | start state, attempt to go north)\*P(start state). We have  
   P(end in a, no wall | attempt to go north) = 0.5\*(0.7\*0.2 + 0.1\*0.2 + 0.1\*0.2 + 0.05\*0.9) + 0.25\*(0.7\*0.9) = 0.27  
   P(end in b, no wall | attempt to go north) = 0.125\*(0.7\*0.2 + 0.1\*0.2 + 0.05\*0.2 + 0.05\*0.9) + 0.25\*(0.1\*0.9) = 0.049375  
   P(end in c, no wall | attempt to go north) = 0.5\*(0.05\*0.9) + 0.25\*(0.05\*0.9 + 0.05\*0.2) + 0.125\*(0.1\*0.9) + 0.125\*(0.1\*0.9) = 0.05875  
   P(end in d, no wall | attempt to go north) = same as for b, by symmetry  
     
   P(no wall | attempt to go north) is the sum of these probabilities, equal to 0.4275. Thus, our resulting belief state is  
   P(a) = 0.27/0.4275 = 0.6316  
   P(b) = 0.049375 / 0.4275 = 0.1155  
   P(c) = 0.05875/0.4275 = 0.1374  
   P(d) = 0.049375 / 0.4275 = 0.1155

## Coding portion

See solution file

# Exercise 7

See solution file

Part 3:

RMSE = 11594.7, Standard Deviation = 11626.8

Since the RMSE is so close to the standard deviation, our model did not do much better than just uniformly predicting the average. So a linear model does not work well here.

The underlined features had the largest absolute valued weights, and thus were the most important.

|  |  |
| --- | --- |
| -2.489103354 | kw\_min\_min |
| -1.778465623 | kw\_avg\_min |
| 0.3203558357 | kw\_max\_avg |
| -0.300757970 | kw\_avg\_avg |
| 0.2723686491 | kw\_max\_min |
| 0.1752703233 | num\_hrefs |
| -0.158175770 | num\_self\_hrefs |
| 0.1581083867 | kw\_min\_avg |
| 0.1016983755 | data\_channel\_is\_lifestyle |
| 0.1010609276 | self\_reference\_avg\_sharess |
| 0.0975773566 | data\_channel\_is\_tech |
| 0.0948516933 | abs\_title\_subjectivity |
| 0.0928313121 | is\_weekend |
| 0.0927966336 | self\_reference\_min\_shares |
| 0.0923656677 | weekday\_is\_monday |
| 0.0919024862 | weekday\_is\_saturday |
| 0.0914995752 | LDA\_03 |
| 0.0912512014 | LDA\_00 |
| 0.0907743662 | LDA\_04 |
| 0.0906044821 | min\_positive\_polarity |
| 0.0885161501 | weekday\_is\_sunday |
| 0.0879602322 | global\_subjectivity |
| 0.0877267364 | avg\_positive\_polarity |
| 0.0876606181 | max\_positive\_polarity |
| 0.0875736816 | global\_rate\_negative\_words |
| 0.0872578575 | global\_sentiment\_polarity |
| 0.0872040739 | global\_rate\_positive\_words |
| 0.0871441333 | min\_negative\_polarity |
| 0.0868476853 | rate\_negative\_words |
| 0.0866901266 | abs\_title\_sentiment\_polarity |
| 0.0855219652 | weekday\_is\_friday |
| 0.0854297711 | n\_non\_stop\_unique\_tokens |
| 0.0853247104 | weekday\_is\_wednesday |
| 0.0852146306 | data\_channel\_is\_socmed |
| 0.0850249547 | n\_unique\_tokens |
| 0.0846668277 | avg\_negative\_polarity |
| 0.0832714243 | weekday\_is\_thursday |
| 0.0829714817 | n\_non\_stop\_words |
| 0.082915057 | num\_imgs |
| 0.0828132567 | n\_tokens\_title |
| 0.0827224869 | title\_subjectivity |
| 0.0819297304 | rate\_positive\_words |
| 0.0816141768 | title\_sentiment\_polarity |
| 0.0813487196 | max\_negative\_polarity |
| 0.0807475256 | weekday\_is\_tuesday |
| 0.080079564 | LDA\_02 |
| 0.0789331652 | num\_keywords |
| 0.0788688579 | LDA\_01 |
| 0.078303628 | data\_channel\_is\_world |
| 0.0760345077 | data\_channel\_is\_entertainment |
| 0.0733746202 | num\_videos |
| 0.0728959932 | data\_channel\_is\_bus |
| 0.0670116303 | average\_token\_length |
| 0.0515521454 | kw\_avg\_max |
| -0.047689028 | n\_tokens\_content |
| -0.042992404 | kw\_min\_max |
| -0.018792893 | kw\_max\_max |
| -0.002492351 | self\_reference\_max\_shares |

# Extra Credit

See solution file.

The 5 most informative features are underlined. kw\_avg and kw\_max\_avg are both among the highest weighted features in the regression model as well.

|  |  |
| --- | --- |
| kw\_avg\_avg | 0.0481668154 |
| kw\_max\_avg | 0.0407445256 |
| self\_reference\_avg\_sharess | 0.0364717726 |
| self\_reference\_min\_shares | 0.032855987 |
| kw\_min\_avg | 0.0287579271 |
| self\_reference\_max\_shares | 0.0273170973 |
| LDA\_02 | 0.0228462433 |
| is\_weekend | 0.0211356477 |
| data\_channel\_is\_world | 0.0205087922 |
| data\_channel\_is\_socmed | 0.0115511258 |
| LDA\_04 | 0.0112952553 |
| LDA\_03 | 0.0112360174 |
| data\_channel\_is\_entertainment | 0.0112314832 |
| weekday\_is\_saturday | 0.0110105637 |
| num\_imgs | 0.0108386465 |
| data\_channel\_is\_tech | 0.0104379609 |
| num\_hrefs | 0.0104336982 |
| global\_subjectivity | 0.0095223499 |
| weekday\_is\_sunday | 0.0093118428 |
| kw\_min\_max | 0.0081093647 |
| kw\_avg\_min | 0.0075834653 |
| min\_positive\_polarity | 0.0074837535 |
| global\_sentiment\_polarity | 0.0069874843 |
| rate\_negative\_words | 0.0062980183 |
| kw\_max\_min | 0.0058661127 |
| LDA\_00 | 0.0057451433 |
| rate\_positive\_words | 0.0056195439 |
| kw\_avg\_max | 0.0055904759 |
| global\_rate\_positive\_words | 0.0055265295 |
| kw\_min\_min | 0.0054860307 |
| LDA\_01 | 0.0054480799 |
| n\_non\_stop\_unique\_tokens | 0.0053006361 |
| n\_unique\_tokens | 0.0051197001 |
| num\_keywords | 0.0043388274 |
| n\_tokens\_content | 0.0040326715 |
| n\_non\_stop\_words | 0.0033488833 |
| num\_self\_hrefs | 0.0030148946 |
| max\_positive\_polarity | 0.0028532595 |
| avg\_positive\_polarity | 0.0028198772 |
| num\_videos | 0.0027812828 |
| kw\_max\_max | 0.0026143862 |
| average\_token\_length | 0.0025463686 |
| title\_subjectivity | 0.0024348899 |
| title\_sentiment\_polarity | 0.0022412308 |
| weekday\_is\_wednesday | 0.0019180813 |
| n\_tokens\_title | 0.0017338433 |
| abs\_title\_sentiment\_polarity | 0.001728744 |
| avg\_negative\_polarity | 0.001559337 |
| data\_channel\_is\_bus | 0.001545074 |
| data\_channel\_is\_lifestyle | 0.001503473 |
| global\_rate\_negative\_words | 0.001433056 |
| max\_negative\_polarity | 0.00143222 |
| weekday\_is\_tuesday | 0.001291873 |
| weekday\_is\_thursday | 0.000901700 |
| min\_negative\_polarity | 0.000792066 |
| weekday\_is\_monday | 0.000580186 |
| weekday\_is\_friday | 0.000341771 |
| abs\_title\_subjectivity | 0.000155938 |