Solutions to Assignment 1:

1 Linear Algebra

1.1 Basic Operations

1. 14
2. 0
3. $$\begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix}$$
4. $\sqrt{5}$
5. $$\begin{bmatrix} 6 \\ 5 \\ 7 \end{bmatrix}$$
6. $$\begin{bmatrix} 19 \end{bmatrix}$$

1.2 Matrix Algebra Rules

1. false, because although the values are the same, the LHS returns a matrix while the RHS returns a value
2. false, because although the values are the same, the LHS returns a matrix while the RHS returns a value
3. false
4. false
5. false
6. true
7. false
8. true
9. true
10. false

2 Probability

2.1 Rules of probability

1. $5
2. 0.55
3. 0.2475

2.2 Bayes Rule and Conditional Probability

1. 0.010096
2. false positives
3. 0.00960777
4. No
5. find the feature to split on that would give you the best probability on the drug test

3 Calculus

3.1 One-variable derivatives

1. $f'(x) = 6x - 2$.
2. $f'(x) = 1 - 2x$
3. $f'(x) = 1 - \frac{\exp(-x)}{p(x)}$.

3.2 Multi-variable derivatives

1. $< 2x_1 + \exp(x_1 + 2x_2), 2\exp(x_1 + 2x_2) >$
2. $< \frac{\exp(x_1)}{Z}, \frac{\exp(x_2)}{Z}, \frac{\exp(x_3)}{Z} >$
3. $< a_1, a_2, a_3 >$
4. $< 2x_1 - x_2, 2x_2 - x_1 >$
5. $< x_1, x_2, x_3, ..., x_d >$

3.3 Optimization

1. $f(x) = 4.25$
2. $f(x) = 0.25$
3. $f(x) = 0$
4. $x = 0, 1$
5. $f(x_1, x_2) = 1$
6. $x_1 = 0, x_2 = 0$

4 Algorithms and Data Structures Review

4.1 Trees

1. 6
2. 5

4.2 Common Runtimes (assuming worse run time

1. $O(n)$
2. $O(\log n)$
3. $O(n)$
4. $O(d)$
5. $O(d^2)$

4.3 Running times of code

1. $O(n)$
2. $O(n)$
3. $O(1)$
4. $O(n)$
5. $O(n^2)$

5 Data Exploration

5.1 Summary Statistics

1. maximum = 4.862, minimum = 0.352, median = 1.177, mean = 1.324625, mode = 0.77
2. $5\% = 0.466, 25\% = 0.72, 50\% = 1.177, 75\% = 1.8145, 95\% = 2.646$
3. highest mean and variance = WtdILI, lowest mean and variance = Pac
4. No, because the values are quantitative and the current mode just appears the most. If we want to give an accurate "common" value, we can use the median because most of the values are similar to it.

5.2 Data Visualization

1. Plot C: histogram showing only the columns (regions) of the dataset
2. Plot D: histogram that displays ALL the values
3. Plot B: Boxplot that has an x-axis in weeks
4. Plot A: y-axis shows the illness percentage
5. Plot F: scatterplot that has a better linear distribution which means better correlation
6. Plot E: scatterplot where points show a slight linear distribution which means its less correlated

6 Decision Trees

6.1 Splitting rule

Equality-based splitting rule would make sense for categorical features. For example, there was a feature that determined whether or not you were male, then it would make more sense to use an equality based split rather than a threshhold split

6.2 Comparig Implementations

If both the error rate and the depth are the same for both approaches, the implementations should be the same. Since the curve is the same, the implementations for the condition each stump splits upon should be the same. There may be the algorithm to iterate through the dataset and implementation can vary but they should be the same if the iteration approach is the same.

6.3 Cost of Fitting Decision Trees

The cost offitting trees in terms of n, m, d is $O(nd\log(n)\log(m))$ as it would take $O(nd\log(n))$ and since there are m nodes, it should take $O(\log(m))$ to compute the tree