Week 4 Homework

Dalton Rothenberger

1.
    - (a) - ff 00 ab 3e 12 b3 –> ff00ab3e,00ab3e12, ab3e12b3
    - (b) - The feature space was hex n-grams that were deemed most relevant based on information gain. The label space was malicious and not malicious.
    - (c) - I understand what information gain returns but I do not understand how the calculation results in that. The TFIDF went over my head, especially how it can use continuous values. How would that even work programmatically.
    - (d) - I was surpised that just by looking at hex n-grams they could determine whether or not a program was malicious. I didn't even know AI/ML could be applied to such a problem.
    - (e) - The information gain function looks at all the n-grams and calculates how much info is gained by a given n-gram. The less frequent n-grams provide more info while the n-grams that appear in every executable provide less. Entropy is embedded in the information gain formula. Information gain is the difference in entropy before and after you know something.
    - (f) - The prior is the proportion of the training data belonging to the given class.
    - (g) - C is the class with the highest probability from the Naïve Bayes calculation. The max function would return the maximum probability calculated.
    - (h) - The authors state stability may be the reason why the boasted SVMs were inconclusive.
    - (i)
        1. The gain ratio is typically used for selecting splitting nodes.
        2. The KL-divergence helps measure how much information is lost due to approximation.
        3. KL-divergence is related to entropy because the calculation of KL-divergence has the entropy of the system as its second term.
2. Naive Bayes uses the following loss function:

    $$1 - P(c|x)P(c)$$

    The function penalizes misclassification. It assigns the smallest loss to the solution that has the greatest number of correct classifications. If that equation returns 0 that means that it predicts 100% correct.
3. Naïve Bayes has a problem with 0s because it uses multiplication in its calculation. If any value in Naïve Bayes is 0 then everything becomes 0 due to the fact that 0 multiplied by any number is 0. Logistic Regression does not have this problem because it uses addition which means when 0 is added nothing happens.