1. Training, Testing, and Validation
   1. What percentage of our dataset did we use for:
      1. Training?
      2. Testing?
      3. Validation?
   2. Explain how training, testing, and validation datasets are used, and why separating our data in this way is important.
2. Record the accuracies for the models that you fit in this activity. Add rows if needed, but please compare a minimum of five models.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features used:** | | | | | | |
| **Number** | **Logistic or SVM** | **Kernel (if SVM)** | **Degree (if polynomial)** | **C-value (if SVM)** | **Training Accuracy** | **Testing**  **Accuracy** |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |

1. Model Selection:
   1. Which model do you expect to have the highest bias, and why?
   2. Rank your models in terms of the error on the training dataset, and explain how this could be used to assess bias.
   3. Are you surprised by which model has the highest error on the training dataset, and why?
   4. Which model do you expect to have the highest variance, and why?
   5. Rank your models in terms of the error on the testing dataset, and explain how this could be used to assess variance.
   6. Are you surprised by which model has the highest error on the testing dataset, and why?
2. Final Model
   1. Which model did you choose to move forward with, and why?
   2. Report the accuracy of this model on the validation dataset.
   3. Report and interpret the confusion matrix from the validation dataset.
   4. Interpret this model biologically.