Dissociating Performance Tiers on Memory Tasks Across the Lifespan of the Rat

An evaluation of various machine learning models to analyze cognitive aptitude

Adele Kapellusch
INFO 521

December 11, 2017

1 Introduction

There is abundant and consistent evidence showing that animals exhibit age-associated memory deficits over their lifespan. What is not well-understood is the variation in how animals age, especially in studies involving inbred animals. A large-scale study is now being conducted to investigate the interaction of many "memory systems" – the hippocampus, prefrontal cortex, perirhinal cortex,and gut-brain interactions – across three different age groups of rats. This work is being accomplished through cognitive memory tasks, brain imaging, histological surveys, and other biological measures. The aim is to pinpoint behavioral and/or biological markers that are associated with better cognitive aging. Considering the volume of data that has been (and will be) collected for this study, it is important to establish a data analysis pipeline that can be used to find interesting patterns and successfully address the questions at hand.

Machine learning techniques are being widely used in many fields, including cognitive science and the neurosciences because of its applications in dealing with large amounts of data. Machine learning goes beyond statistical inference by also addressing questions about what computational architectures and algorithms can be used to most effectively capture and store data, how multiple learning subtasks can be orchestrated in a larger system, and questions of computational tractability. In this report, I show preliminary results of a shallow and broad excavation of a portion of the behavioral component of this project. I compare different approaches as well as different possible models to represent the cognitive aptitude of this group of rats as well as make predictions about other animals. I hope this can serve as a "proof-of-concept" for future data analysis of the entirety of this study's data.

2 Methods

All data was collected by members of the laboratory of Dr. Carol Barnes and analyses described here were implemented using the Sci-kit learn Python library. Here I give brief descriptions of the theory behind each model used.

2.1 Cognitive Task Data Collection

307 animals were tested on 4 cognitive tasks (2 of which are included in this report) to get an indirect measure of various types of memory. Of these 307 animals, 93 were young (6 months old), 103 were middle-aged (15 months), and 111 were old (23 months).

2.1.1 Morris Water Navigation Task: Spatial Memory

The Morris Water Navigation Task is a common cognitive task used to test spatial learning and memory in an animal model. A rat is placed at a random location along the edge of a pool filled with opaque water. The rat has 60 seconds to locate a hidden platform submerged approximately 2 cm beneath the water level. The animal is tested for 12 trials and the platform remains in the same place. The starting position is randomly varied across trials which requires the animal to use the visual cues in the room to navigate within the pool, thus giving us a measure of spatial learning. The rat's position is tracked while he swims and the corrected integrated path length

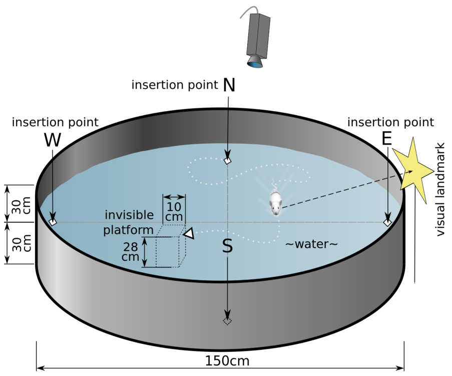


Figure 1: Diagram of Morris Water Navigation Task [2]

(CIPL) is calculated using the sum of the distances from the target at each acquired point during the trial with a correction factor to compensate for the initial distance from the goal for each starting location [1]. This is the measurement that will be used to quantify performance. A low CIPL value means that the animal took a direct path to the platform whereas a high CIPL value means that the animal took a meandering path before discovering the location of the platform.

2.1.2 Modified Water Navigation Task: Working Memory

The water navigation task can be modified to test working memory. The rat is placed in the water and must swim to find the hidden platform, just as in the spatial version. The difference, however, is that the rat must then rest for a delay period where they must hold the position of the platform in working memory after which they must swim again to find the platform. Rats are tested at three different delays: 30 seconds, 30 minutes, and 2 hours. The CIPL is again the behavioral measurement calculated from the tracking data.

2.2 Regression

2.2.1 Ordinary Least Squares

This model assumes a linear relationship between the input variables (x) and output variable (y). y can be calculated from a linear combination of of the x's. [5]

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \quad (1)$$

The linear equation assigns one scale factor (i.e. coefficient) to each input value or column. The model is fit so that the coefficients $(\beta_0, \beta_1, \ldots \beta_p)$ minimize the mean squared error (MSE) between the observed responses in the dataset and the responses predicted by the linear approximation.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \quad (2)$$

The fit linear model can tell us if there is a relationship between the response and the predictors and can allow us to make predictions for the response given some predictors.

2.2.2 Lasso

Lasso (Least Absolute Selection Shrinkage Operator) is a regularized form of linear regression. Regularization penalizes complex/extreme representations by adding a term that represents the sizes of the coefficients to the loss function. By imposing this constraint, you are essentially preventing overfitting. Lasso has an absolute value regularization term so that you are not only punishing models with high coefficient values, but also setting the coefficient values to zero if they are not relevant. This shrinkage allows for better interpretation of the model and identifies which features are most strongly associated with the response variables.

$$\min_{w} \frac{1}{2n_{samples}} \|Xw - y\|_2^2 + \alpha\|w\|_1 \tag{3}$$

The penalty is the added $\alpha\|w\|_1$ term, where α is a constant and $\|w\|_1$ is the l_1-norm of the parameter vector. [4]

2.3 Classification

Various different classifiers were fit using the behavioral data and asked to classify whether an animal was young, middle-aged or old.

2.3.1 K-Nearest Neighbors

K-Nearest Neighbors is a popular and general-use method for classifying data non-probabilistically. The model classifies a new point by finding the K training points that are closest to the new point (x_{new}). t_{new} is set to be the majority class amongst neighbors. [6] Here, I choose K using cross-validation (see section 2.6.1).

2.3.2 Support Vector Machines

The support vector classifier is a generalization of the maximal margin classifier; it classifies based on constructing a separating hyperplane and determining on which side the hyperplane the point lies. It constructs this by finding the hyperplane that is farthest from the training observations. This is done by computing the perpendicular distance from each observation and the smallest distance is known as the *margin*. To have an optimal support vector classifier, you must maximize the margin. The support vector machine (SVM) is an extension of the support vector classifier that results from enlarging the feature space using kernels [5]. This allows us to accommodate a non-linear boundary between each class.

Because this case requires a non-binary classification (we have three classes:young, middle-aged, and old), this method needs to be extended to the "one-against-one" approach.[7] This is done by constructing $\binom{K}{2}$ SVMs, each of which trains data from two classes.

2.3.3 Naive Bayes

Naive Bayes is a probabilistic classifier, meaning that it predicts a probability distribution over a set of classes rather than just using a distance measure to return what class a new point belongs to. This model uses Bayes' theorem with a strong assumption of independence between all the features (this is the *Naive* component). Bayes' theorem states the following:

$$P(y|x_1,...,x_n) = \frac{P(y)P(x_1,...,x_n|y)}{P(x_1,...,x_n)} \tag{4}$$

with the assumption:

$$P(x_i|y, x_1, ...x_{i-1}, x_{i+1}, ...x_n) = P(x_i|y) \tag{5}$$

and since $P(x_1, ...x_n)$ is constant given the input, the expression can be simplified to this for all i:

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y) \prod_{i=1}^{n} P(x_i|y) \tag{6}$$

We can then use a Maximum A Posteriori (MAP) estimation to estimate $P(y)$ and $P(x_i|y)$. There are different types of Naive Bayes: Gaussian, Multinomial, and Bernoulli. These different classifiers only differ in the assumption they make about the distribution of $P(x_i|y)$.

2.4 Clustering

2.4.1 K-Means

K-means clustering is an approach to partitioning data into K distinct clusters. The within-cluster variation for cluster C_k is a measure $W(C_k)$ of the amount by which the observations within a cluster differ from each other. The goal is to minimize the within-cluster variation as much as possible. The most common method of doing this is minimizing the squared Euclidean distance.

$$\underset{C_1,\ldots,C_K}{\text{minimize}}\Big\{\sum_{k=1}^{K}\frac{1}{|C_k|}\sum_{i,i'\epsilon C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'k})^2\Big\} \tag{7}$$

The approach is as follows: randomly assign a number from 1 to K to each of the observations which will serve as initial cluster assignments. We then iterate: for each of the clusters, compute the cluster center and assign each observation to the cluster whose centroid is closest. We will iterate this process until the cluster assignments stop changing.[5]

2.5 Model Evaluation

A critical step in analysis is determining if the predictive modeling techniques are appropriate for the questions we want to ask. It is very important to evaluate a predictive model by seeing how it will generalize to an independent dataset (data that it has never encountered before).

2.5.1 Supervised Learning Model Evaluation

One of the most common performance measures of a supervised learning model for classification is prediction accuracy (ACC). Which is essentially the proportion of responses that were correctly predicted. I will be using this measure for classification performance and MSE for regression performance.

$$ACC = 1 - \frac{1}{n}\sum_{i=1}^{n}L(\hat{y}_i, y_i) \tag{8}$$

$$L(\hat{y}_i, y_i) := \begin{cases} 0 & \text{if} \quad \hat{y}_i = y_i \\ 1 & \text{if} \quad \hat{y}_i \neq y_i \end{cases} \tag{9}$$

The next step is to use the chosen performance measure and do cross validation for model evaluation. k-Fold Cross-validation gives the opportunity for all samples to get tested. The data is split into k *folds* and on each iteration, one fold is used as the validation set and the rest is used as training data. After iterating k times, the average performance is calculated and this is our measure of how well the model did on unseen data. [6]

2.5.2 Unsupervised Learning Model Evaluation

If the ground truth is not known, Silhouette Coefficient can be used to evaluate the model using the model itself [3]. A higher silhouette coefficient means that the result has better defined clusters.

$$s = \frac{b-a}{max(a,b)} \tag{10}$$

a: mean distance between a sample and all points in the same class
b: mean distance between the sample and all other points in the next nearest cluster
Calinski-Harabaz Index can also be used to evaluate clustering. For k clusters, the Calinski-Harabaz score s is given as the ratio of the between-clusters dispersion mean and the within-cluster dispersion [3]

$$s(k) = \frac{Tr(B_k)}{Tr(W_k)} \times \frac{N-k}{k-1} \tag{11}$$

$$W_k = \sum_{q=1}^{k}\sum_{x\epsilon c_q}(x-c_q)(x-c_q)^\top \tag{12}$$

$$B_k = \sum_q n_q (c_q - c)(c_q - c)^\top \tag{13}$$

Where B_k is the between group dispersion matrix and W_k is the within-cluster dispersion matrix. Again, a higher Calinski-Harabaz score means better defined clusters. I use this and the Silhouette Coefficient to give me a measure of how well my unsupervised clustering does in grouping the data.

3 Results

3.1 Comparing learning rates of young, middle-aged, and old animals

A simple linear model was fit using water maze data to compare the different learning rates for each age group. Young animals had the fastest learning rate, then middle-aged animals, and then old animals (see figure 2). These findings are not surprising, but in order to determine if a linear assumption of this data was appropriate, the data was fit using models of polynomial degree 1,2,3,4, and 5 (see figure 3 and 4) and looked at the relative mean squared error for each. After

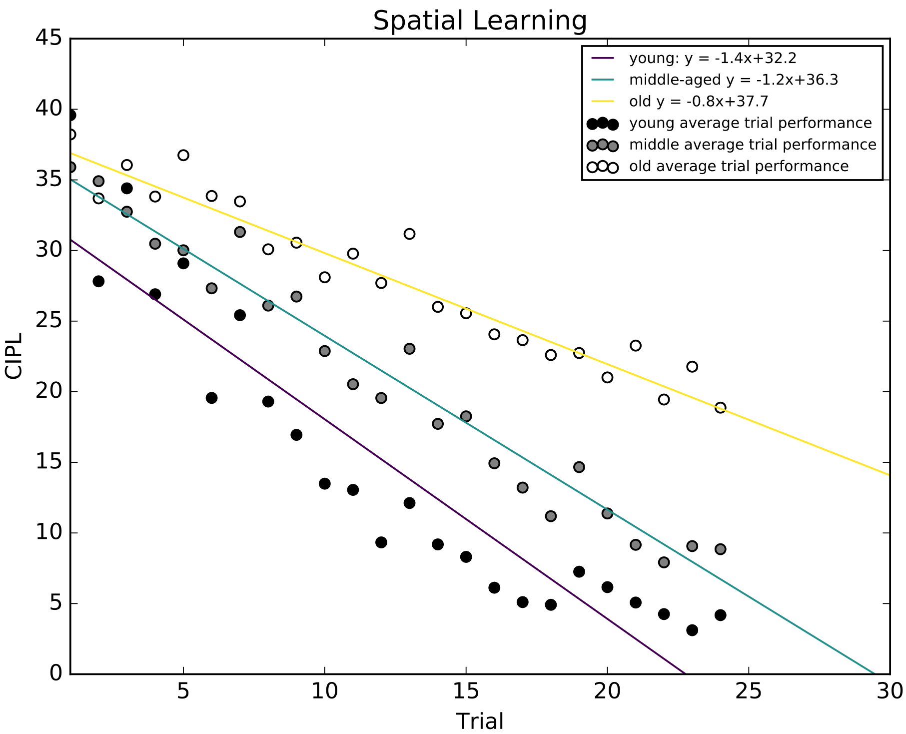
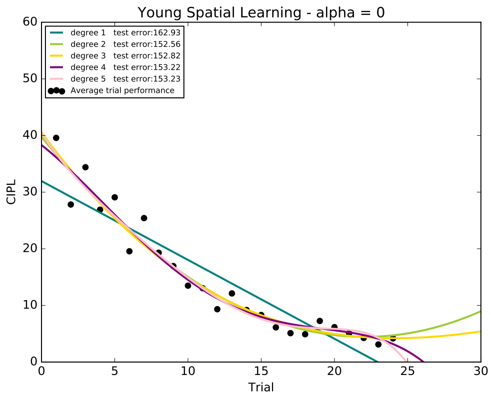
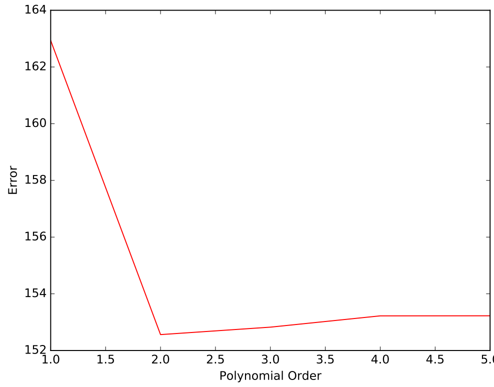


Figure 2: Linear models plotted for each age group. The points are the average trial performance for each age group. Young animals have the highest slope learning rate, then middle-aged, then old animals.

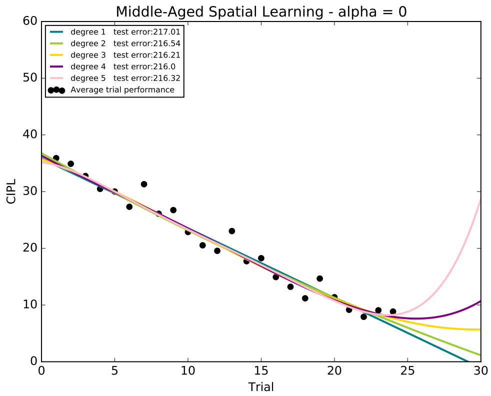
regularizing, it seemed that all three datasets were best described by models of polynomial degree 2. These are useful to look at because one can think about the slope at each trial representing how much information the average animal within an age group learned on that trial; or how much improvement is seen on a given trial.



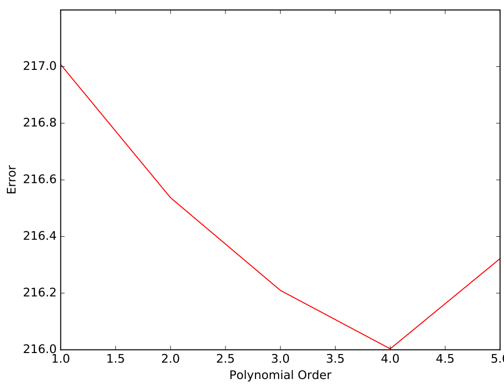
(a)



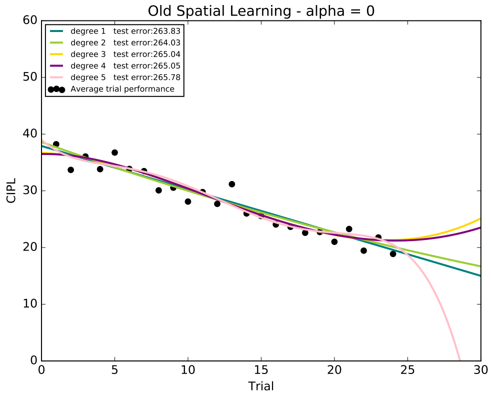
(b)



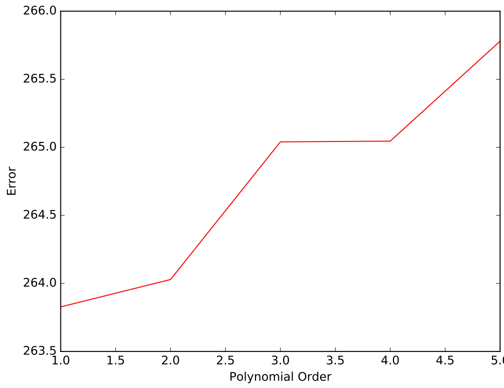
(c)



(d)



(e)



(f)

Figure 3: Unregularized linear models fit for polynomial degrees 1 - 5. Young rats' data and model error comparison (a and b), middle-aged rats' data and model error comparison (c and d), and old-aged rats' data and model error comparison (e and f).

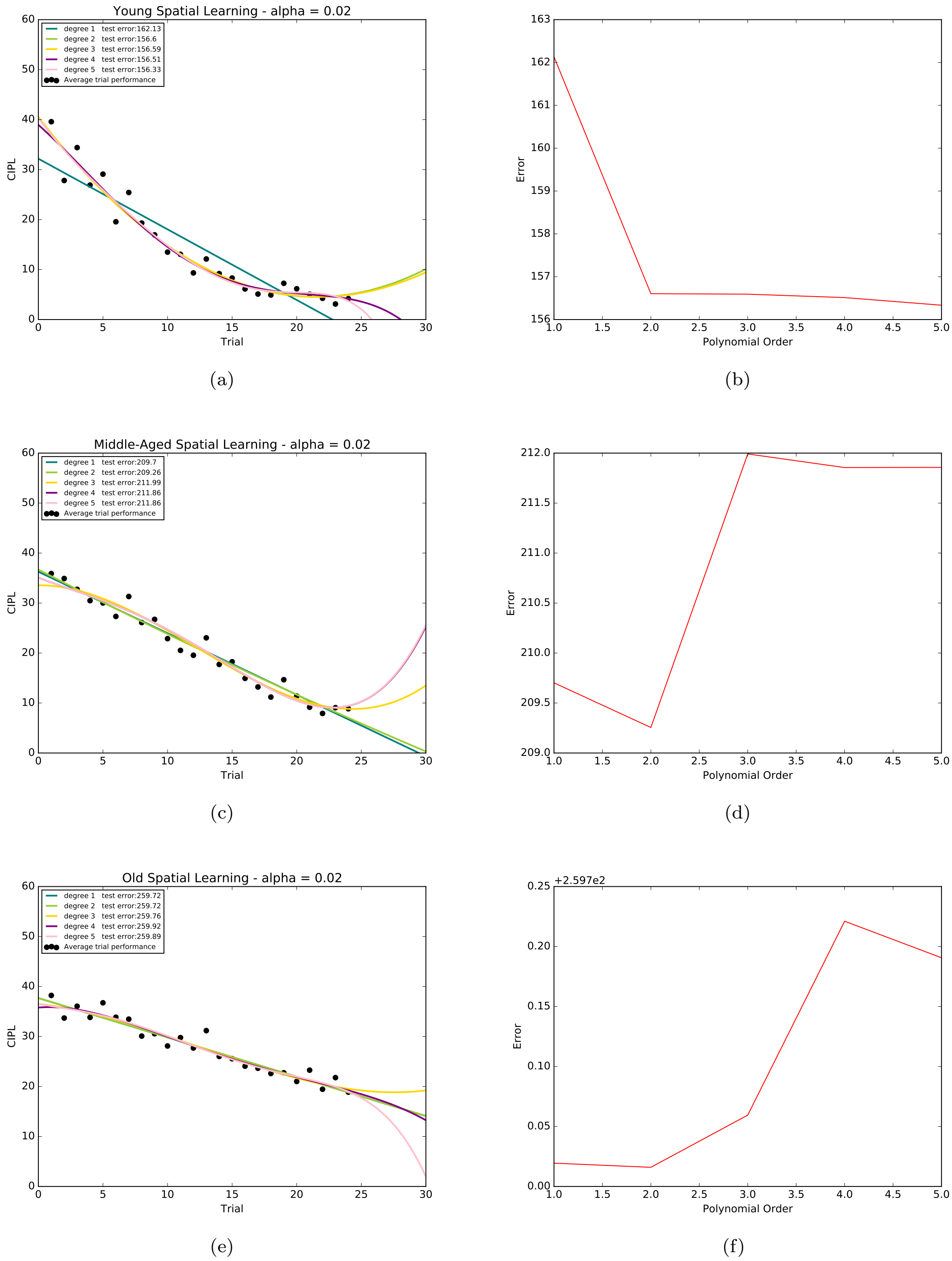


Figure 4: Regularized (alpha = 0.02) linear models fit for polynomial degrees 1 - 5. Young rats' data and model error comparison (a and b), middle-aged rats' data and model error comparison (c and d), and old-aged rats' data and model error comparison (e and f). It seems that for this value of alpha, a regularized model of degree 2 is best for all three sets of data.

3.2 How does cognitive abilities deteriorate with age?

Next, I did a regression analysis of age and task performance to see the relationship. There is a relatively high correlation between spatial memory and the age of a rat (see figure 5). However, for all three delay periods for the working memory task show a weak correlation between performance and age (see figure 6). Performance on both of these tasks were used as features in a final linear model to see if this could capture both types of learning over the age of the animal (see figure 7). These models can make predictions about how an animal of an age group that was not tested (e.g. 9 months) would do on a spatial or working memory task.

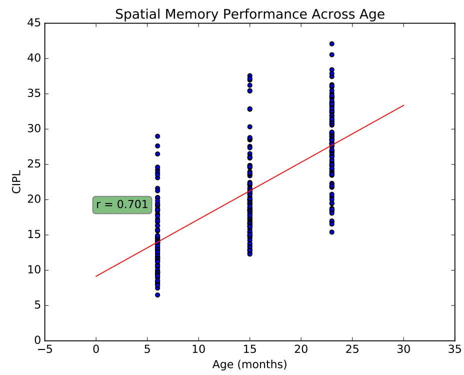
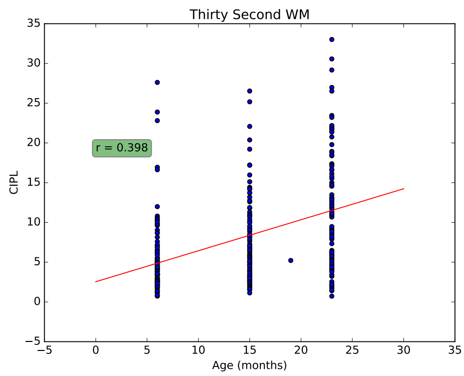
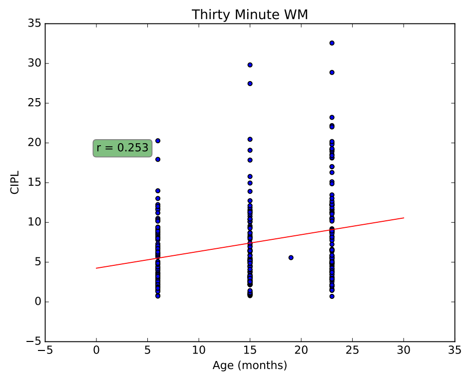


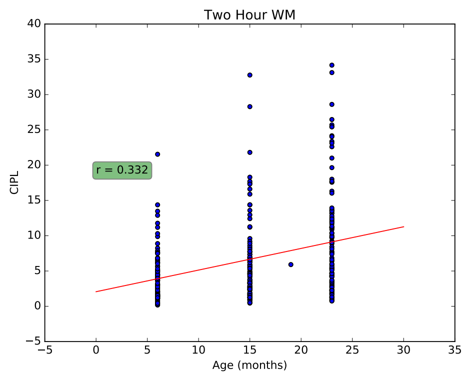
Figure 5: Average (across all trials) performance on the water maze task across age groups. Linear regression shows a strong relationship between spatial learning and age (pearson correlation coefficent = 0.701).



(a)

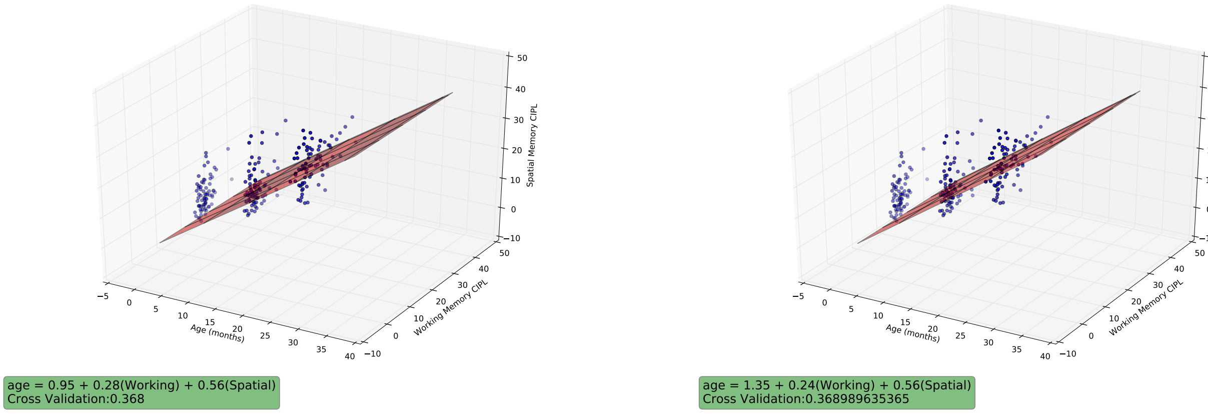


(b)



(c)

Figure 6: Average (across all trials) working memory performance for thirty second delay (a), thirty minute delay(b), and two hour delay(c). Regression shows a very weak relationship between working memory performance and age.



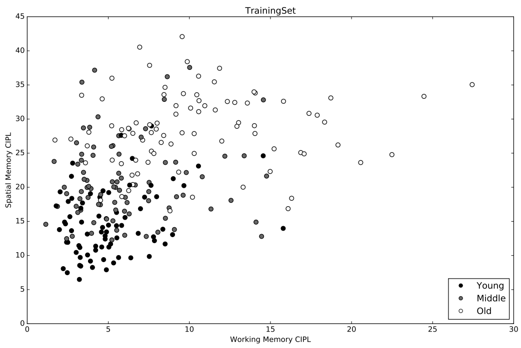
(a) (b)

Figure 7: Linear model fit to working memory and spatial memory performance. The training data and resulting hyperplane is plotted. (a) is a ordinary least squares model and (b) is using lasso. There is not much difference in the results.

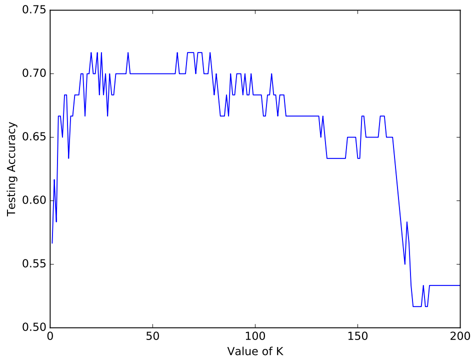
3.3 Classifying age groups based on cognitive task performance

I next trained models to take performance on the working memory and spatial tasks and classify an animal as being young, middle-aged, or old. This can give us insight into how successfully an animal is aging compared to the population of laboratory rats.

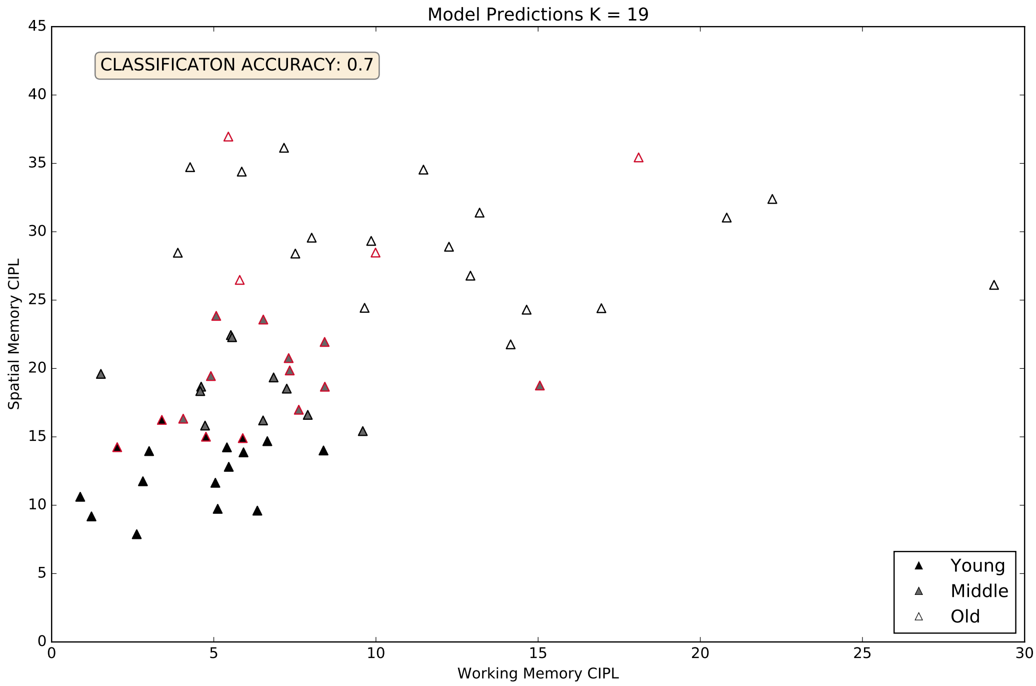
This was first done using K-Nearest Neighbors (see figure 8), then Naive Bayes (figure 9), and finally SVM (figure 10). Based on 10-fold cross validation scores, KNN was a better classifier for this data.



(a)



(b)



(c)

Figure 8: KNN to classify age from working memory performance and spatial memory performance. Training data shows the true classifications (a) and (b) shows the testing accuracy for different values of K. For this data, the best K was determined to be 19. In (c) we see the predictions made by the classifier on the testing set. Incorrect predictions are outlined in red. The classifier successfully classified an animal approximately 70 percent of the time.

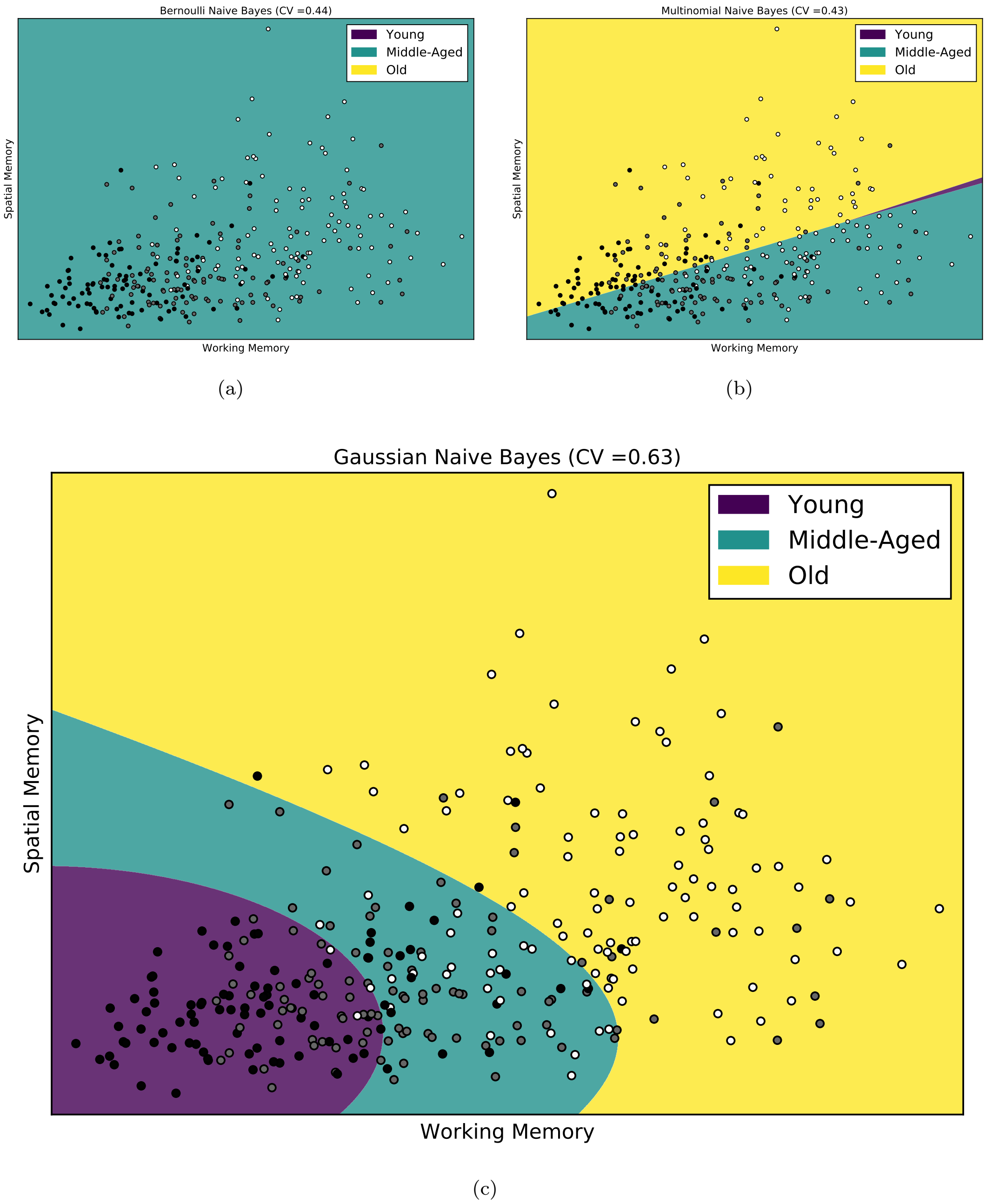


Figure 9: Results of a Naive Bayes classifier with bernoulli assumption (a), multinomial assumption(b), and gaussian assumption (c). The bernoulli classifier does not do a good job classifying the data (this makes sense since the data cannot be distributed in this way) and it classifies all animals as middle-aged. The multinomial classifier also does a poor job. The Gaussian classifier classifies rats fairly well, but not as good as KNN.

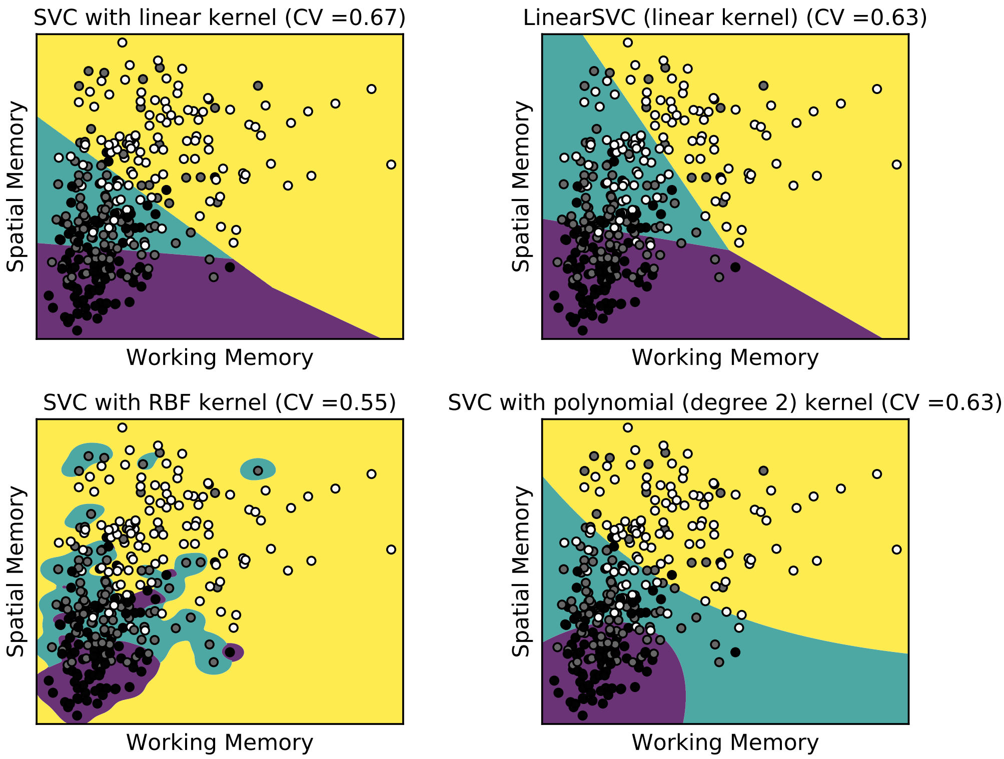


Figure 10: Results of SVM classifier with different kernels. The linear kernel does the best at classifying while the RBF kernel does not capture the global patterns of the data and does not do a good job classifying the data.

3.4 Is performance on a spatial memory task a linear predictor of performance on a working memory task?

One of the questions I wanted to address in this analysis is whether or not a rat's ability to perform well on a spatial-memory-dependent task is an indicator of how well it will perform on a working-memory-dependent task. To answer this question of orthogonality, the data was fit to a linear model (see figure 11). When using all of the data as training data, there was no correlation between task performances. Furthermore, when broken down into age groups, there is a weak to no correlation between performance on both of these tasks. This suggests that spatial learning ability is not a linear predictor of working memory performance.

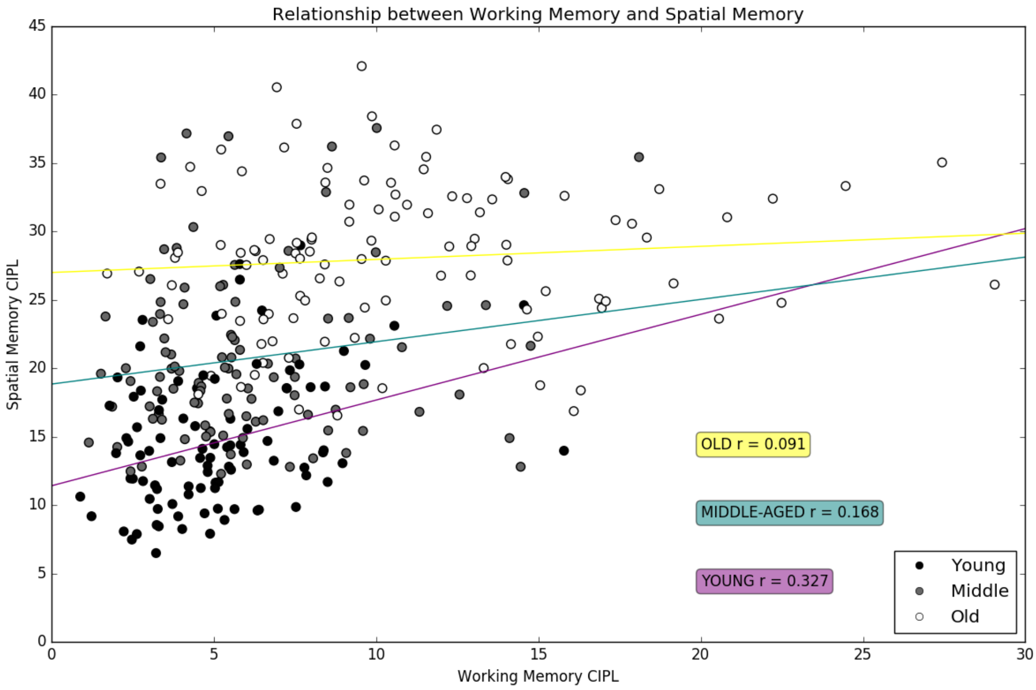
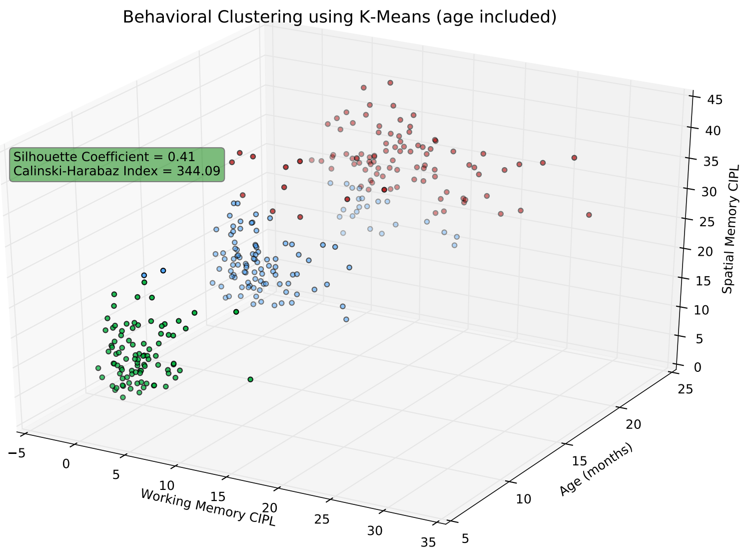


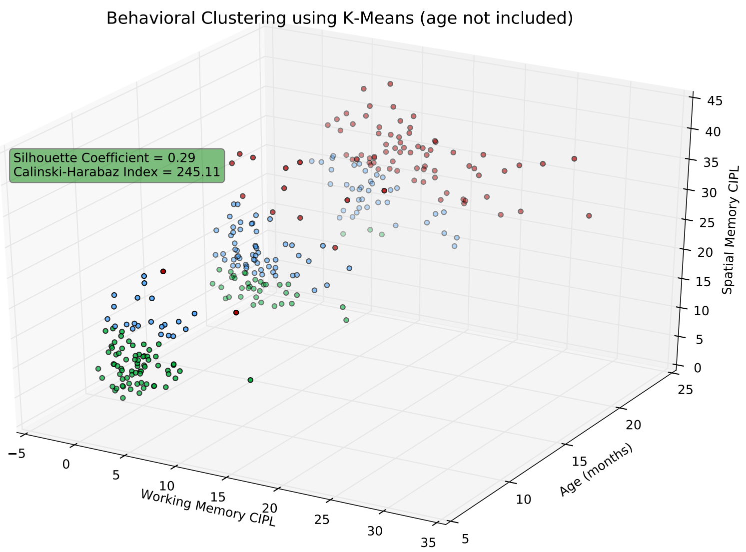
Figure 11: Fit linear models for each age group where Spatial Memory performance is the response. There is no correlation between working memory performance and spatial memory performance in this group of rats.

3.5 Grouping performance levels

Another way the lab is interested in looking at the data is grouping rats into "low-performers", "average-performers", and "high-performers". This was done using KMeans clustering. The model first included three features: working memory performance, spatial memory performance, and age. This resulted in clusters that were too strongly influenced by the age feature. Another model was created which was blind to age and this resulted in an interesting distribution of the different learner groups.

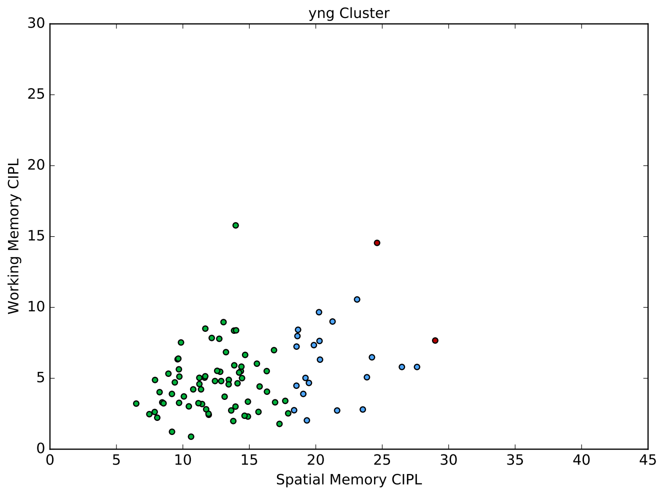


(a)

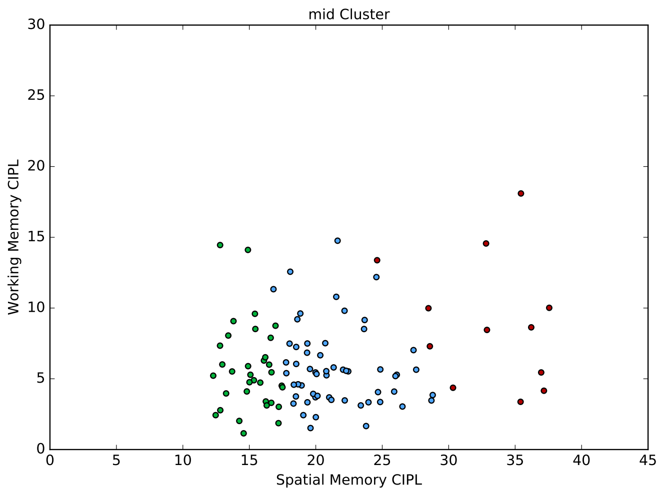


(b)

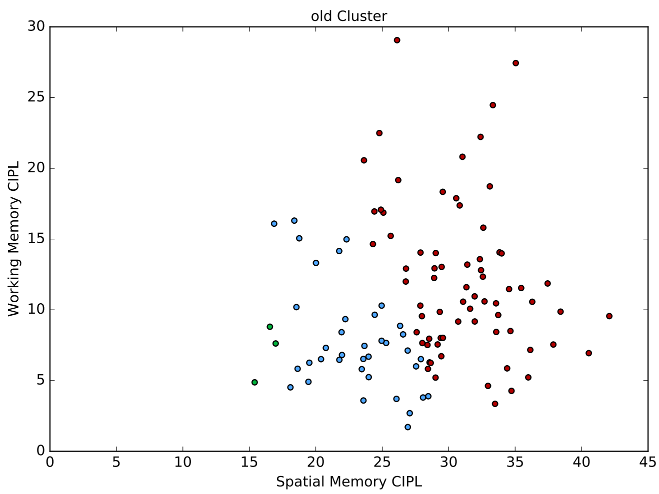
Figure 12: Clustering dataset using KMeans. (a) shows a model that had age as a feature. (b) is a model that was blind to age; it is just plotted in three-dimensions for easier comparison to the first model.



(a)



(b)



(c)

Figure 13: These plots show cross-sections of the clusters from the model shown in figure 12(b). It is interesting to see the proportion of each learner group in each age group.

4 Discussion

This brief look at this behavioral dataset has revealed some interesting patterns.

Learning rates on the Morris Water Navigation Task are fastest for young rats, then middle-aged rats, then old rats and these learning rates seem to be non-linear. Learning rates were best described using a regularized quadratic. It also seems that spatial memory deteriorates more with age while there was less of a difference in performance on this particular working memory task between the three age groups. It is valuable, however, to use the model that has both tasks as features to predict how an animal of some age will do on the tasks.

Classification by KNN yielded the best results for this dataset. It had a 70 percent accuracy, which may not seem great, but is expected because of the large variation in performance within age groups. It will be useful to use the rats that were classified incorrectly to make conclusions about how they are aging– e.g. if a middle-aged animal is classified as being old, he may be aging "poorly". It was also clear that how well an animal does on the spatial memory task does not indicate how well they will do on the working memory task. This is consistent with results from human studies.[8]

Grouping the entire population of rats into "high performers", "average performers", and "low performers" shows that young animals have a large proportion of high-performers and not many average and low performers. The middle-aged group has nearly equal proportions of all learner groups while the aged animals have the majority of low-performers. This classification can be useful for classifying animals for use in other studies in the laboratory.

These python scripts and results serve as a proof-of-concept to demonstrate which machine learning algorithms are best to analyze this type of data. In the future I would like to also include data from the other cognitive tasks these animals have completed as well as tissue data and gut bacteria sequencing data. Incorporating these features into the models may yield some exciting new findings correlating different biological markers and what cognitive aptitude tier an animal is a member and how successfully an animal is aging.

5 Code Documentation

Code Repository: https://github.com/adelekap/INFO521FinalProject

All required packages and their versions are in `requirements.txt`. To install the required packages just run:

```
pip install -r requirements.txt
```

Data Files

`CASWatermaze.csv` This csv has the water maze task data.

`CASWorkingMemory.csv` This csv has the working memory task data.

`data.py` This module houses the dataframes required for training and testing models.

Python Scripts

`KMeans.py` Run this to perform performance clustering using KMeans

`KNN.py` Run this to see prediction of ages using KNN

`lasso.py` Run this to fit lasso linear model using performances of both tasks

`learning.py` Fits linear models of different polynomial degrees to the water maze data

`LeastSquares.py` Fits and visualizes linear models for 1 and 2 features

`naiveBayes.py` Fits data for gaussian, multinomial, and bernoulli naive bayes

`SVM.py` Trains SVM classifier and plots contours

References

[1] Gallagher, M. (1993) *Severity of spatial learning impairment in aging: Development of a learning index for performance in the Morris water maze.* Behavioral Neuroscience, 107(4), 618-626.

[2] John, S. (2010) (2) John, S. (2010). MorrisWaterMaze [Digital image]. Retrieved from https://commons.wikimedia.org/wiki/File:MorrisWaterMaze.svg MorrisWaterMaze [Digital image].

[3] Clustering : 2.3.9.5. Silhouette Coefficient. Retrieved December 07, 2017, from http://scikit-learn.org/stable/modules/clustering.html

[4] Regression: 1.1.3 Lasso. Retrieved December 08, 2017, from http://scikit-learn.org/stable/modules/linear_model.html

[5] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017) *An introduction to statistical learning with applications in R.* New York: Springer.

[6] Rogers, S., and Girolami, M. (2017) *A first course in machine learning.* Boca Raton: CRC Press, Taylor and Francis Group, a Chapman and Hall book.

[7] Knerr, S., Personnaz, L., and Dreyfus, G. (1990) Single-layer learning revisited: A stepwise procedure for building and training neural network. . Neurocomputing: algorithms, architectures and applications, 68, 41-50.

[8] Glisky,E., Kony, L. (2008) Do young and older adults rely on different processes in source memory tasks? Journal of Experimental Psychology, 34:4, 809-822.