Student Names: Nate Davis and Alexander Lobo
Collaboration Statement:

Turning in this assignment indicates you have abided by the course Collaboration Policy:

`www.cs.tufts.edu/comp/136/2022s/index.html#collaboration-policy`

Total hours spent: 1

We consulted the following resources:

- Course Website
- UCI Machine Learning Repository

FYI Official instructions for this checkpoint can be found at:

`www.cs.tufts.edu/comp/136/2022s/checkpoint0.html`

Note: the submission length limit for this is one page (not including the cover page). You should not be writing more than a few sentences to a paragraph for each of these pieces.

Team

Team Members: Nate Davis and Alexander Lobo

Dataset

Dataset Access

Access: `https://archive.ics.uci.edu/ml/datasets/Musk+%28Version+2%29`

Description:

The dataset contains measures of distances within different shapes (conformations) of a set of 102 molecules. The study that this data comes from used human experts to judge the smell of each molecule and determine whether it is characterized as ”musk” or ”non-musk”, which makes this a binary classification dataset.

Dataset Properties

Number of instances: There are 6,598 conformations total.

Number of features: There are 166 features (not including the molecule and conformation names).

Model-Related Properties

We plan on using a probabilistic logistic regression model to characterize the data. We intend to use a multi-variate normal prior on the weight vector. We plan to use feature ”class” as the binary target for our probabilistic logistic regression model.