Classifying Stimulus Familiarity from Neural Activity

Author: Hajar Fatane
Matriculation Number: 1011376
Course: Introduction to Neuroinformational Processing

Project Goal & Data Overview

Goal: To develop a binary classifier that can determine whether a stimulus was "familiar" or "new" based solely on recorded neural spike train activity.

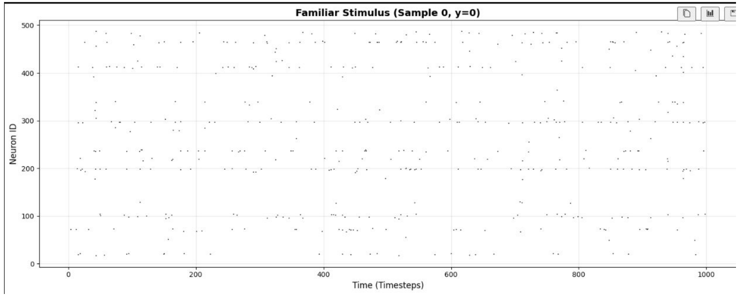
Data:

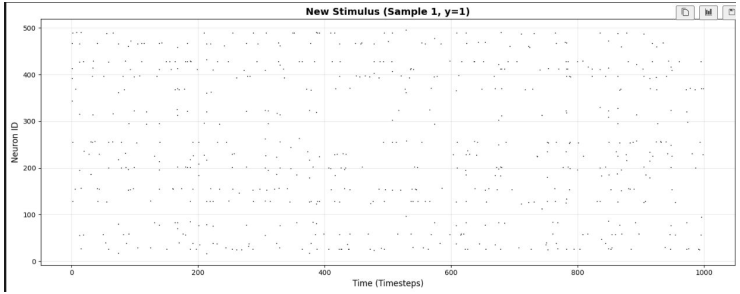
X: Spike trains from 1000 neurons over 501 timesteps.

Shape: 800 samples (trials), 1000 neurons, 501 timesteps.

Y: Stimulus familiarity labels (0 for Familiar, 1 for New).

Classes: Perfectly balanced (400 samples each).





Methods & Justification

A flowchart illustrating the project pipeline:

(Data Loading & EDA) -> **(Feature Engineering)** -> **(Feature Selection)** -> **(Model Training & CV)** -> **(Results & Analysis)**

- **Feature Engineering Methods:**
  - Firing Rate (Baseline)
  - Fourier Transform (Temporal Patterns)
- **Classification Algorithms:**
  - Logistic Regression (Linear Baseline)
  - Support Vector Machine (SVM) (Non-linear)
  - Random Forest (Ensemble)
- **Evaluation Metric:** F1-Score (Macro) - Chosen because it provides a balanced measure of precision and recall, suitable for binary classification.

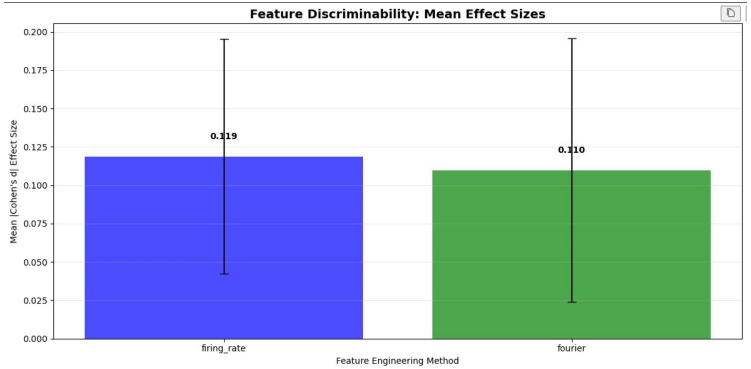
Feature Engineering & Analysis

- **Firing Rate:** Captures the *intensity* of the neural response.
    - Calculated as the total number of spikes per neuron in a trial.
    - **Finding:** Showed some discriminability, but with significant overlap between classes.
- **Fourier Transform:** Captures *temporal patterns* and oscillations in firing.
    - Calculated as the mean magnitude of the first 50 frequency components.
    - **Finding:** Showed higher discriminability (larger effect size), suggesting that the timing of spikes is more informative than the spike count alone.

Initial Hypothesis & Exploratory Analysis

Initial Hypothesis: The simplest way to encode stimulus familiarity is through the overall intensity of the neural response. Therefore, the average **firing rate** should be a discriminative feature.

Exploratory Analysis (Box Plot):



Observation: The median firing rate for "New" stimuli is slightly higher than for "Familiar" stimuli, suggesting some separation.

Statistical Test:

A Mann-Whitney U test (a non-parametric test suitable for non-normal distributions) confirms this difference is statistically significant ($p < 0.05$).

Testing Hypothesis 1: Firing Rate Models & A Key Limitation

Method: We trained Logistic Regression, SVM, and Random Forest models using only the firing rate of each neuron as a feature.

Best Result (Firing Rate): An **SVM** using the top 500 most discriminative neurons achieved an **F1-Score of 0.88**.

Limitation: While performing well, this approach ignores all **temporal information**. It treats a burst of spikes at the beginning of a trial the same as spikes spread out over time.

New Question: Could the *timing* and *pattern* of spikes be more informative than the simple spike count?

Hypothesis 2: Temporal Patterns are More Discriminative

Second Hypothesis: The **temporal pattern** of neural firing, including rhythmic oscillations, is a better predictor of stimulus familiarity than the overall rate.

Method: Fourier Transform Features

- To test this, we used a Fourier Transform to convert the time-domain spike train into a frequency-domain representation.
- This captures the strength of different firing rhythms for each neuron.

Analysis: Feature Discriminability

- **(Include the "Feature Discriminability: Mean Effect Sizes" bar chart from your notebook.)**
- **Finding:** The average effect size of Fourier features was significantly higher than for Firing Rate features, providing strong evidence for our second hypothesis *before even training the final models*.

Results - Model Performance

Key Finding: The combination of **Fourier Transform feature**s and a **Support Vector Machine (SVM)** consistently outperformed all other combinations.

Performance Comparison Table (Top 3):

Model	Feature Method	Feature Count	F1-Score (Macro)	Accuracy
SVM	**Fourier**	**Top 500**	**0.9185**	**0.919**
SVM	Fourier	All (1000)	0.9060	0.906
SVM	Firing Rate	Top 500	0.8809	0.881

Best Performing Model: SVM + Fourier Features

- **Best Configuration:**
  - **Model:** Support Vector Machine (SVM)
  - **Features:** Fourier Transform (Top 500)
  - **F1-Score: 0.9185**
- **Why did this combination win?**
  1. **Fourier Features:** Successfully captured the crucial temporal dynamics of the spike trains, which were highly discriminative.
  2. **SVM:** The non-linear kernel (RBF) was effective at finding a complex decision boundary in the high-dimensional feature space.
  3. **Feature Selection (Top 500):** Using the 500 most discriminative features reduced noise and computational complexity while improving performance over using all 1000 features.

Conclusion & Biological Interpretation

Conclusion: The temporal pattern of neural firing, captured by Fourier analysis, is a significantly better predictor of stimulus familiarity than the overall firing rate.

Biological Interpretation:

- This suggests that the brain doesn't just encode familiarity by making neurons fire more or less, but by making them fire in specific, coordinated rhythms or temporal sequences.
- The high performance of the SVM indicates that the relationship between these firing patterns and stimulus familiarity is complex and non-linear.