

# Predicting Song Popularity: Comparison and Analysis of Machine Learning Methods

BIN PHAM

# Outline

- ▶ Overview
- ▶ The Dataset
- ▶ Methods
- ▶ Results
- ▶ Conclusion

# Overview

- ▶ **Motivation:** The music industry is continuously growing and in order to remain competitive, companies need to be able to predict a song's popularity.
- ▶ **Goals:** Clarify under what conditions which methods perform better. Compare and analyze the advantages and disadvantages of various methods.

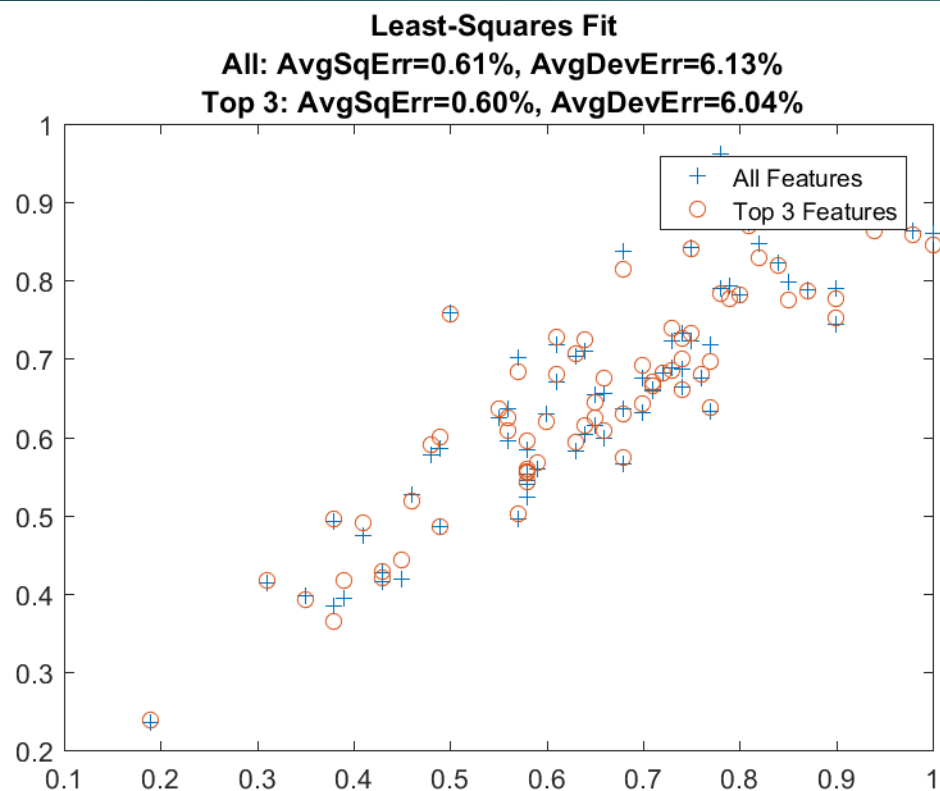
# The Dataset: Personal Sleep Cycle

- ▶ 697 Samples
- ▶ 7 features
  - ▶ Start of Sleep
  - ▶ End of Sleep
  - ▶ Duration of Sleep
  - ▶ Heart Rate
  - ▶ Mood
  - ▶ Activity (Steps)
  - ▶ Sleep Quality

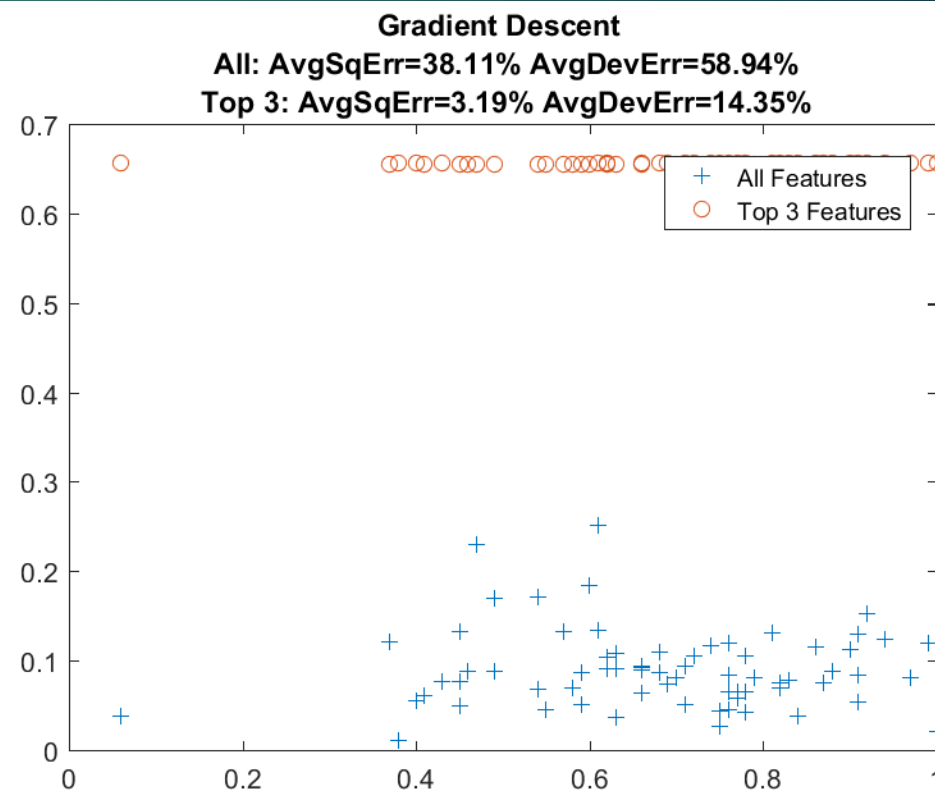
# Methods

- ▶ Linear Regression
  - ▶ Least-Square Fit
  - ▶ Gradient Descent
    - ▶ Regularization
- ▶ Multi-Class Support Vector Machine (multi-SVM)
- ▶ Perceptron Neural Network
  - ▶ Varied 10-1000 neurons

# Results

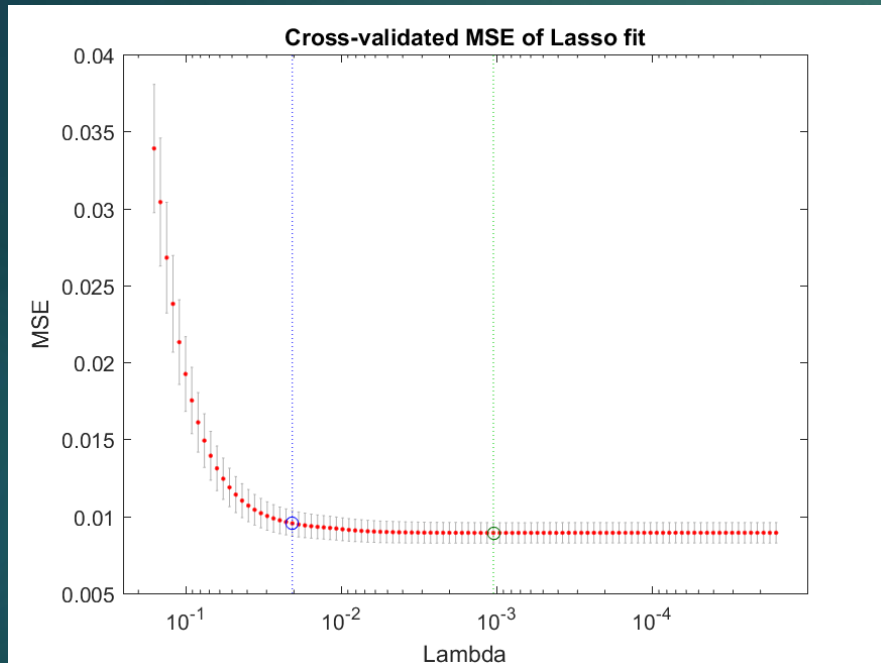


Least-Squares Fit Comparison

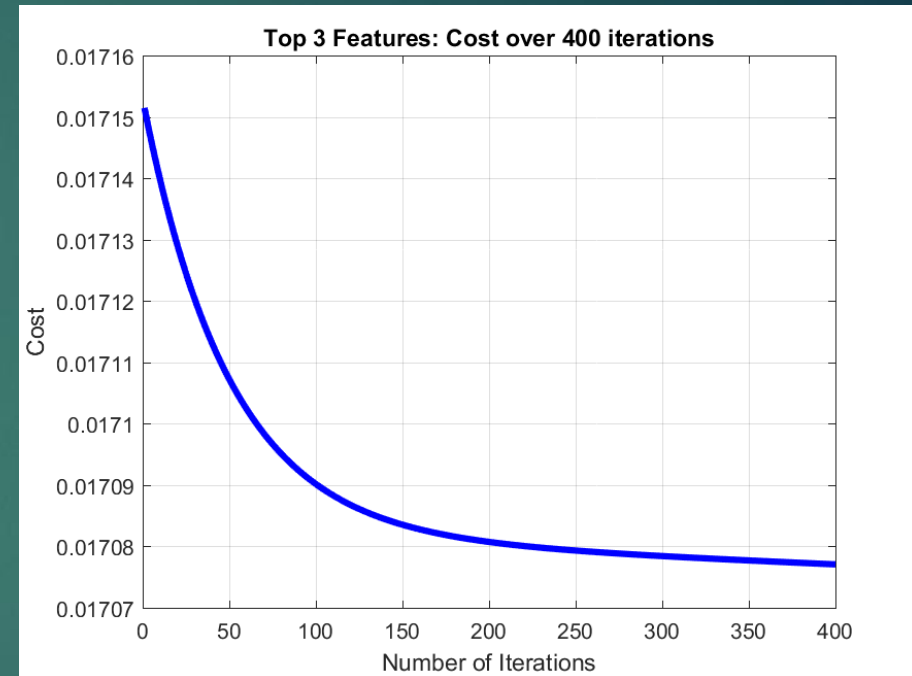


Gradient Descent Comparison

# Results



Gradient Descent w/ Regularization



Iterations of Gradient Descent Cost

# Results

**Confusion Matrix**

Output Class	Target Class	
	0	1
0	0 0.0%	0 0.0%
1	0 0.0%	69 100%

NaN% NaN%  
100% 0.0%

Mult-SVM

**Training Confusion Matrix**

Output Class	Target Class	
	0	1
0	0 0.0%	5 1.0%
1	0 0.0%	483 99.0%

NaN% NaN%  
99.0% 1.0%

**Validation Confusion Matrix**

Output Class	Target Class	
	0	1
0	0 0.0%	1 1.0%
1	0 0.0%	103 99.0%

NaN% NaN%  
99.0% 1.0%

**Test Confusion Matrix**

Output Class	Target Class	
	0	1
0	0 0.0%	0 0.0%
1	0 0.0%	104 100%

NaN% NaN%  
100% 0.0%

**All Confusion Matrix**

Output Class	Target Class	
	0	1
0	0 0.0%	6 0.9%
1	0 0.0%	690 99.1%

NaN% NaN%  
99.1% 0.9%

Perceptron Neural Network



# Conclusion

- ▶ SVM is a great method for small datasets
- ▶ Regularization helps improve prediction accuracy and reduces overfitting of data
- ▶ Least-Square Fit and Gradient Descent perform similarly. Each method has its advantages and disadvantages
- ▶ Deep Learning should be the most accurate and effective method for classifying moods
- ▶ More data from other people is required to definitively determine whether the model is accurate to others