



SeeGull: Pacific Sea Gull Activity Classification and Path Prediction

William Cai¹, Hagop Chinchinian²

¹Department of Mathematics, Stanford University

²Department of Electrical Engineering, Stanford University

Stanford
Computer Science

Abstract

Project goal: Pacific sea gull behavior prediction, including:

- activity classification** based on positional data
- path prediction** based on past positions and weather

Activity Classification
Achieved reasonable results with feature mapping and logistic regression with some dimensionality reduction.

Path prediction
Further work is needed to resolve issues with the LSTM-based path prediction algorithm.

Activity Classification

Task: Given 5 positions, determine class of gull movement.
Class 1: Minor to none
Class 2: Minor, multi-directional
Class 3: Moderate, unidirectional
Class 4: Significant, unidirectional

Data:
Labeled by a group member.
Train data uses positions of 5 gulls.
Test data uses positions of 2 gulls.

Feature maps:

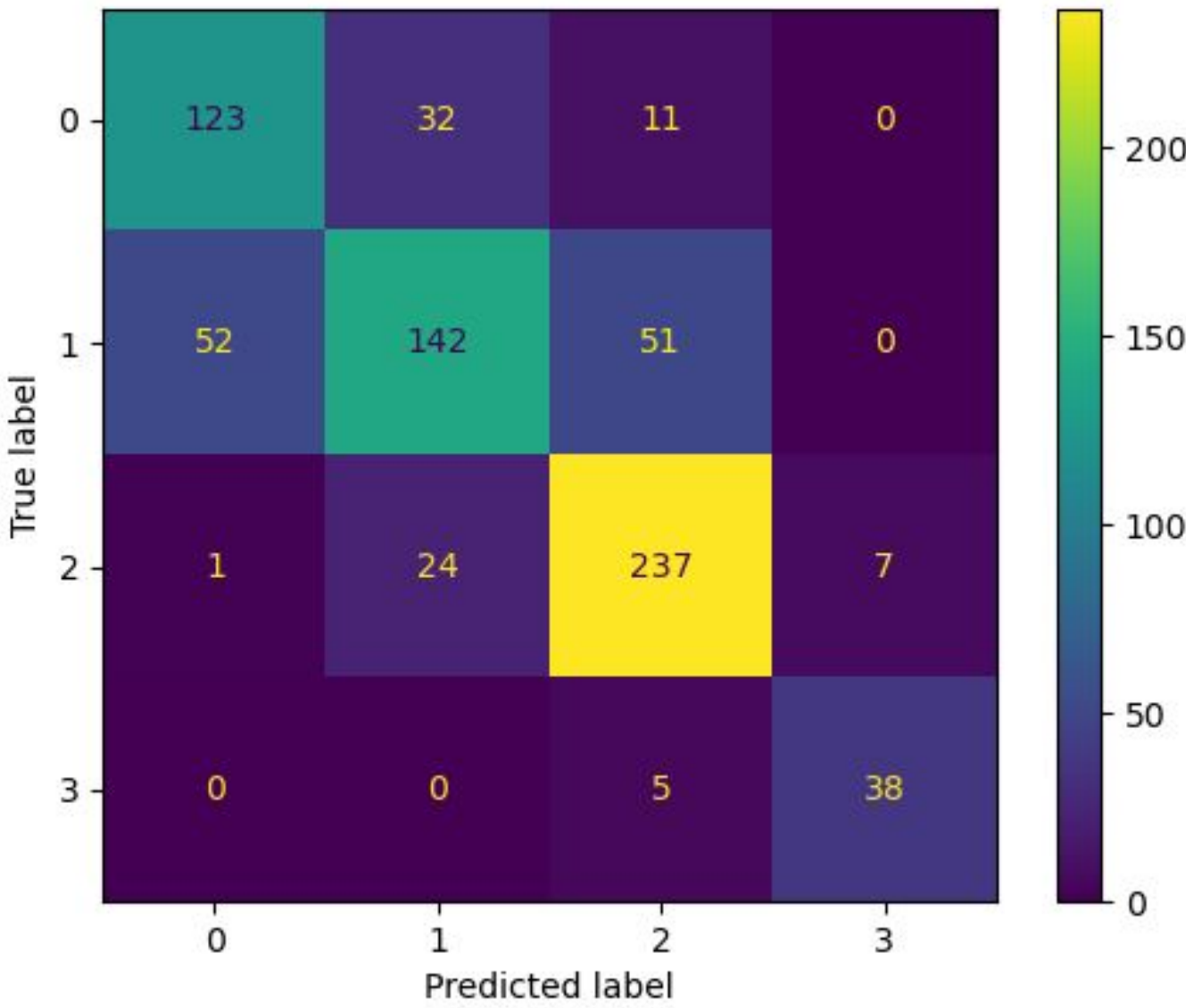
latitude, longitude → distances, directions, standard deviations

Models Considered: Support Vector Machine (SVM), Naive Bayes, Logistic Regression

Dimensionality Reduction: After choosing best model, reduce feature dimensionality with principal component analysis (PCA).

Results: Logistic regression shows the best performance. Logistic regression maintains reasonable performance with number of dimensions reduced to 10 by PCA.

Algorithm	Average F1	Weighted Average F1
SVM, RBF	51%	46%
SVM, poly	59%	55%
SVM, sigmoid	36%	32%
Naive Bayes	67%	62%
Logistic (PCA=19)	77%	75%
Logistic (PCA=9)	66%	66%
Logistic (PCA=10)	76%	74%



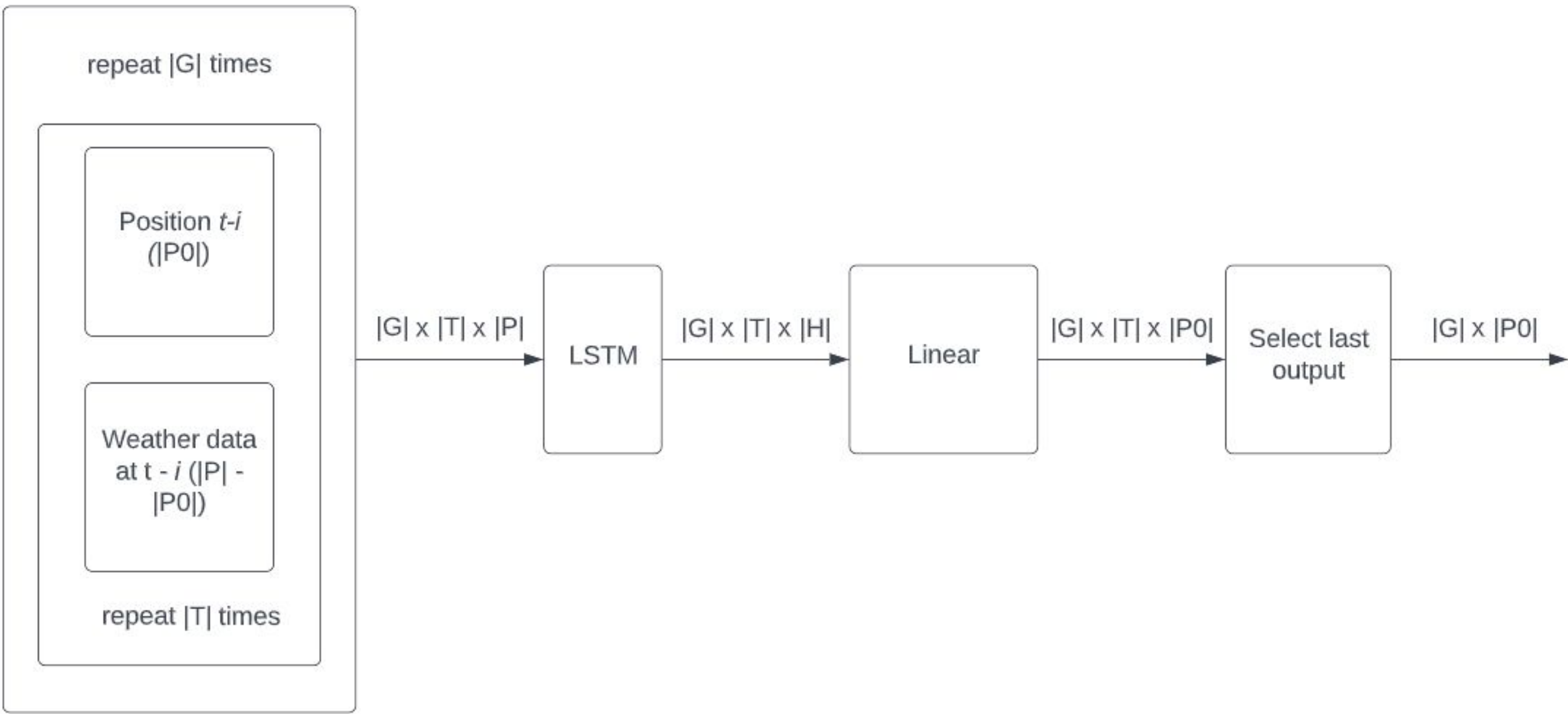
Confusion matrix for logistic regression, PCA-reduced to 10 dimensions

Path Prediction

Task: Given 5 prior locations of the gull as well as weather data across the geographical area at the same respective times, predict the next position of the gull.

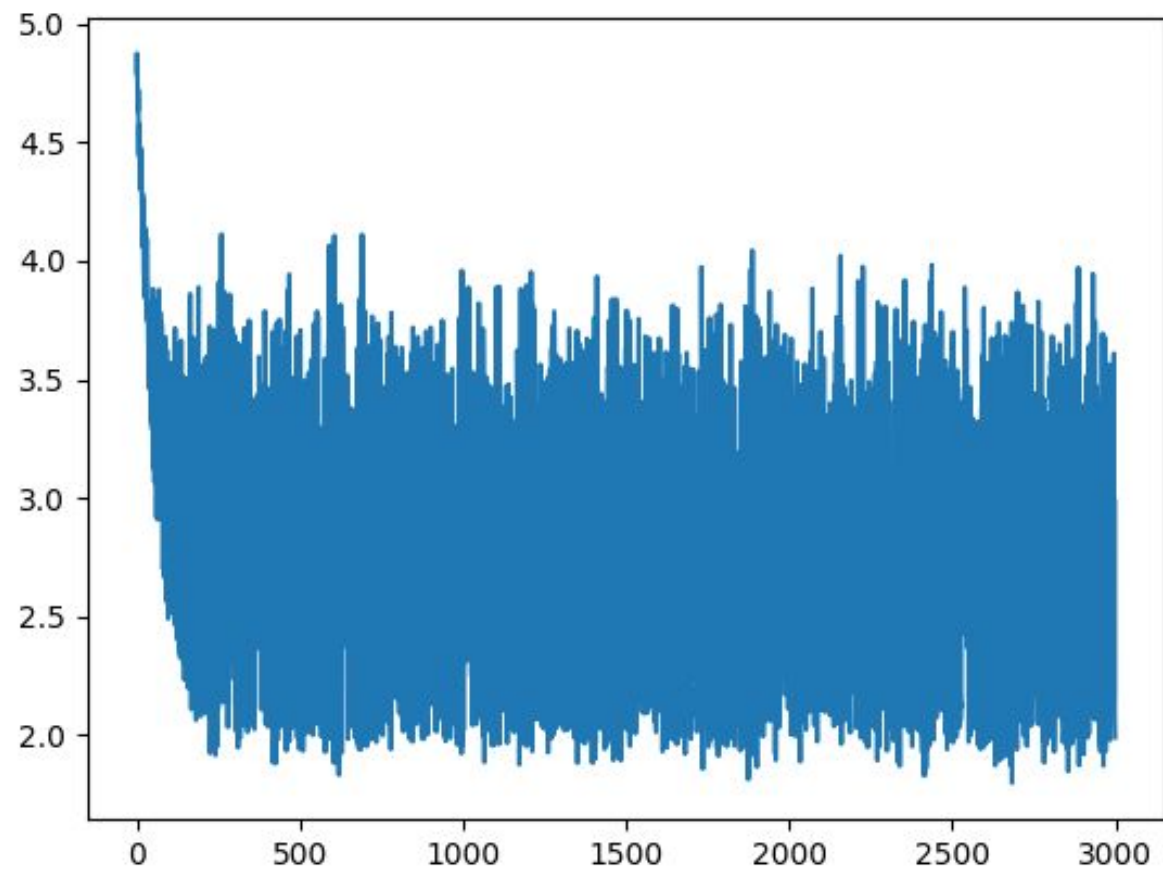
Data:
Train data: first 70% of times.
Test data: Last 25% of times.

Model: Neural network with long short-term memory and linear layers.



Path Prediction Model Architecture

Performance
Model failed to converge on train data and yielded repeated cross-entropy values in the range of 3.3-3.5 on test data.



Training loss as a function of epoch



Stanford
University