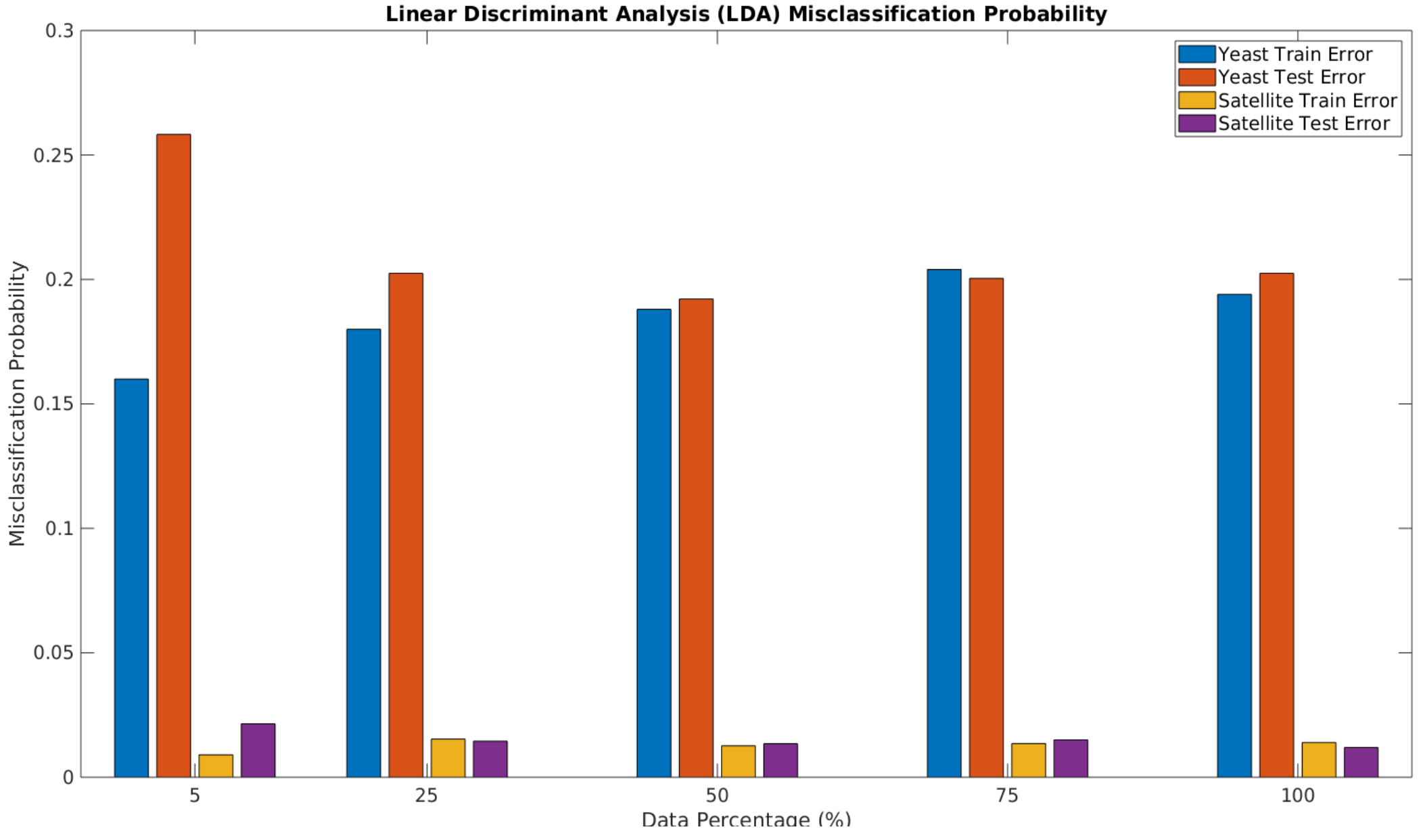
Machine Learning HW #3

Javier Salazar 1001144647

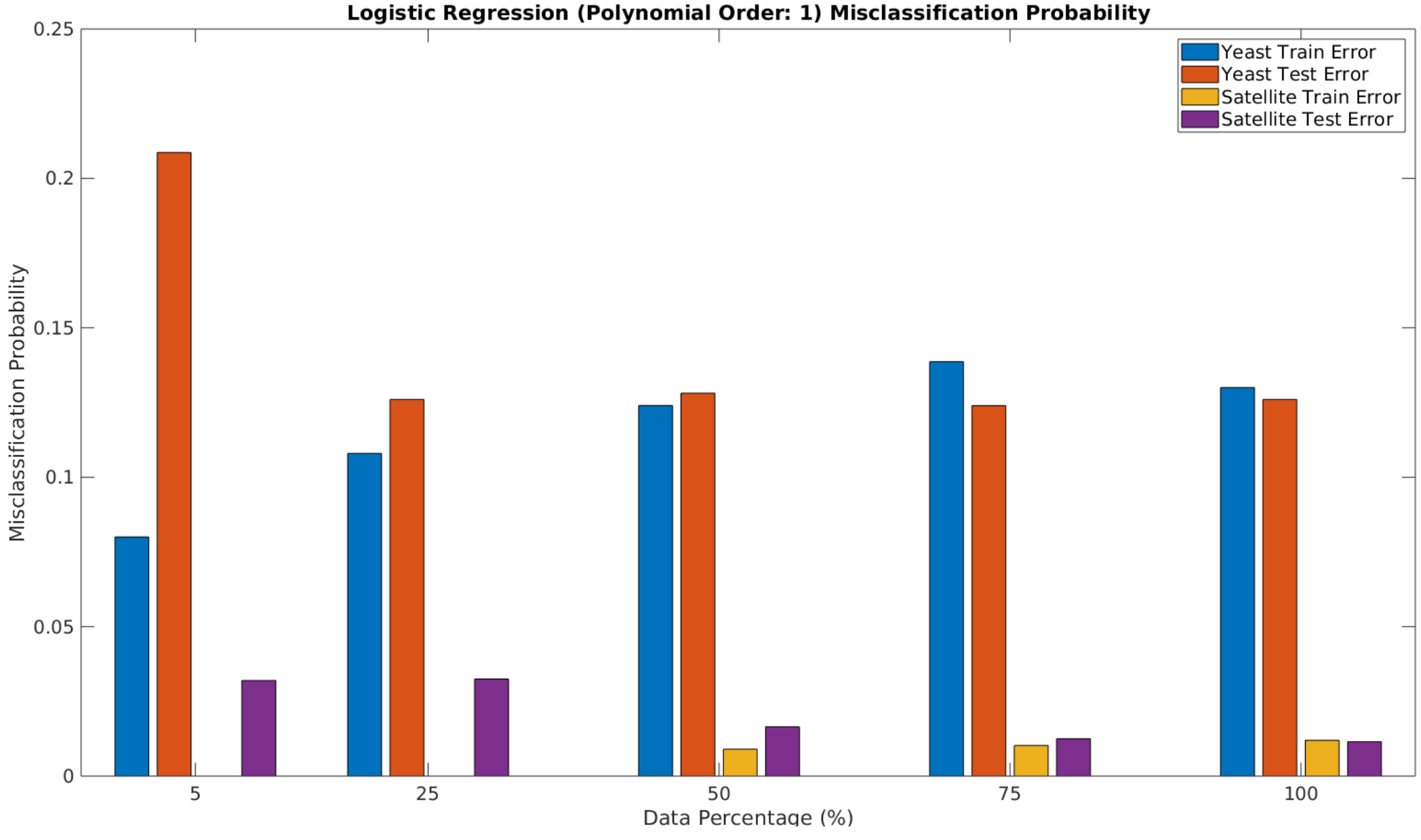
Problem #1 (LDA)



LDA Analysis

- **Very small lambda 1e-12 is used to prevent singular covariance matrix when using 5% of training data**
- **Results stabilize after using 50% of training data or around 500 data points**
- **Satellite training data consists of 5 times the amount of points as Yeast dataset therefore we expect lower misclassification probability**

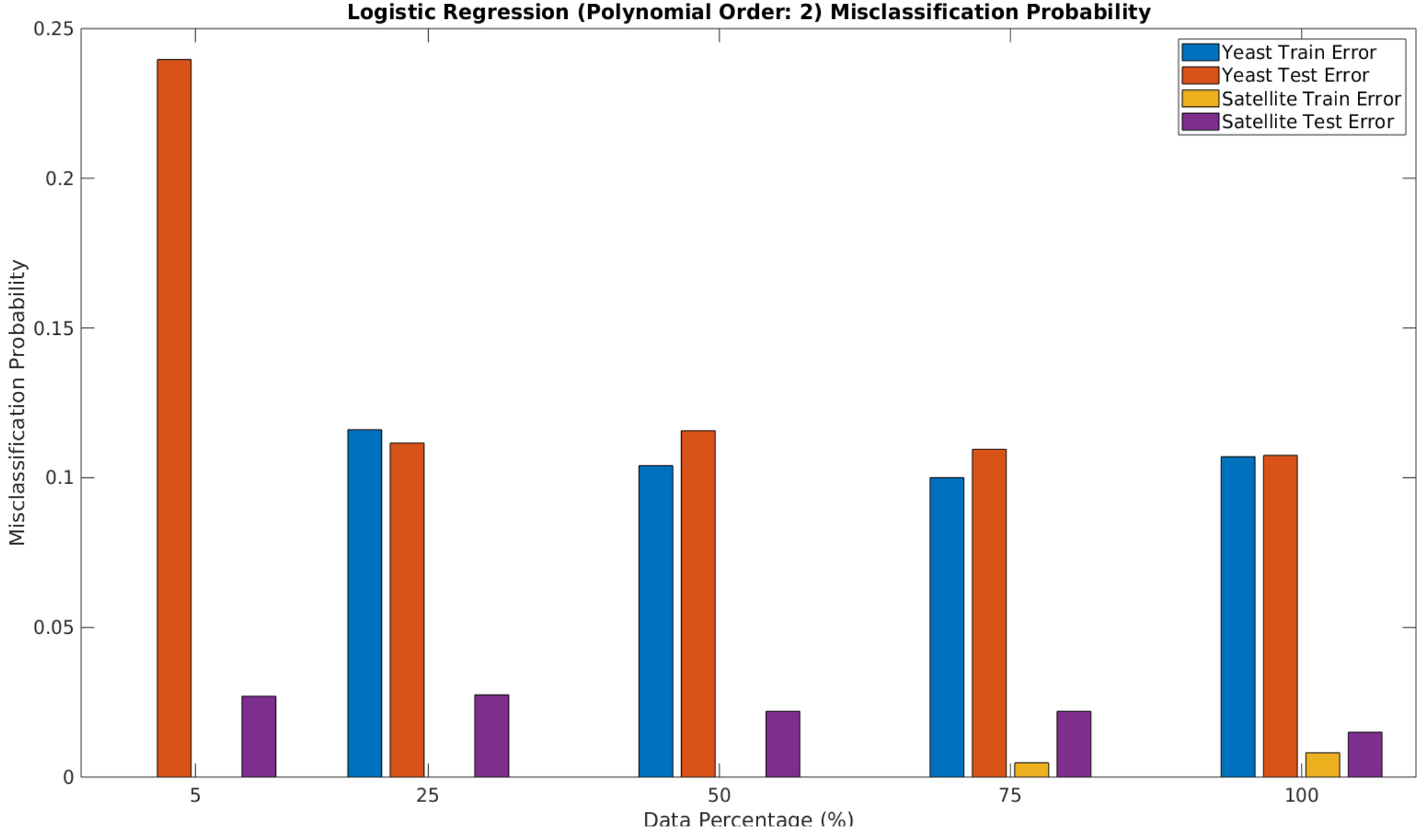
Logistic Regression – Order: 1



Logistic Regression – Order 1 - Analysis

- **Lambda 0.1 used to ensure convergence in Newton-Ralphson method**
- **Stable results after 50% training data used except for numerical instability issue seen at 75% training data where training error slightly bigger than test error**
- **Model fits training data perfectly when we have low amount of training data but will still have testing error as expected**

Logistic Regression – Order: 2



Logistic Regression – Order 2 - Analysis

- **Order 2 does an even better job at fitting training data when lower amount of data is used but will still have test error**
- **Misclassification starts to stabilize after 25% of training data is used**
- **Lambda 0.1 used to prevent singular matrix during Newton-Ralphson method**

Overall Analysis

Method	Training Misclassification (Yeast / Satellite)	Testing Misclassification (Yeast / Satellite)
LDA – 100% Data	0.194 / 0.01398	0.2025 / 0.012
Logistic (Order 1) – 100% Data	0.13 / 0.01195	0.126 / 0.0115
Logistic (Order 2) – 100% Data	0.107 / 0.008117	0.1074 / 0.015

- From above, lowest training error is using order 2 LR followed by order 1 LR and then LDA. This is exactly what we expect to see for both datasets. For the testing error, we get the lowest on order 1 LR surprisingly. Order 2 LR performed the worst (relatively). However, as seen from the bar graphs, order 2 LR was better at lowering testing error when there was less training data (i.e. 25% 50% 75%) relative to order 1 LR.