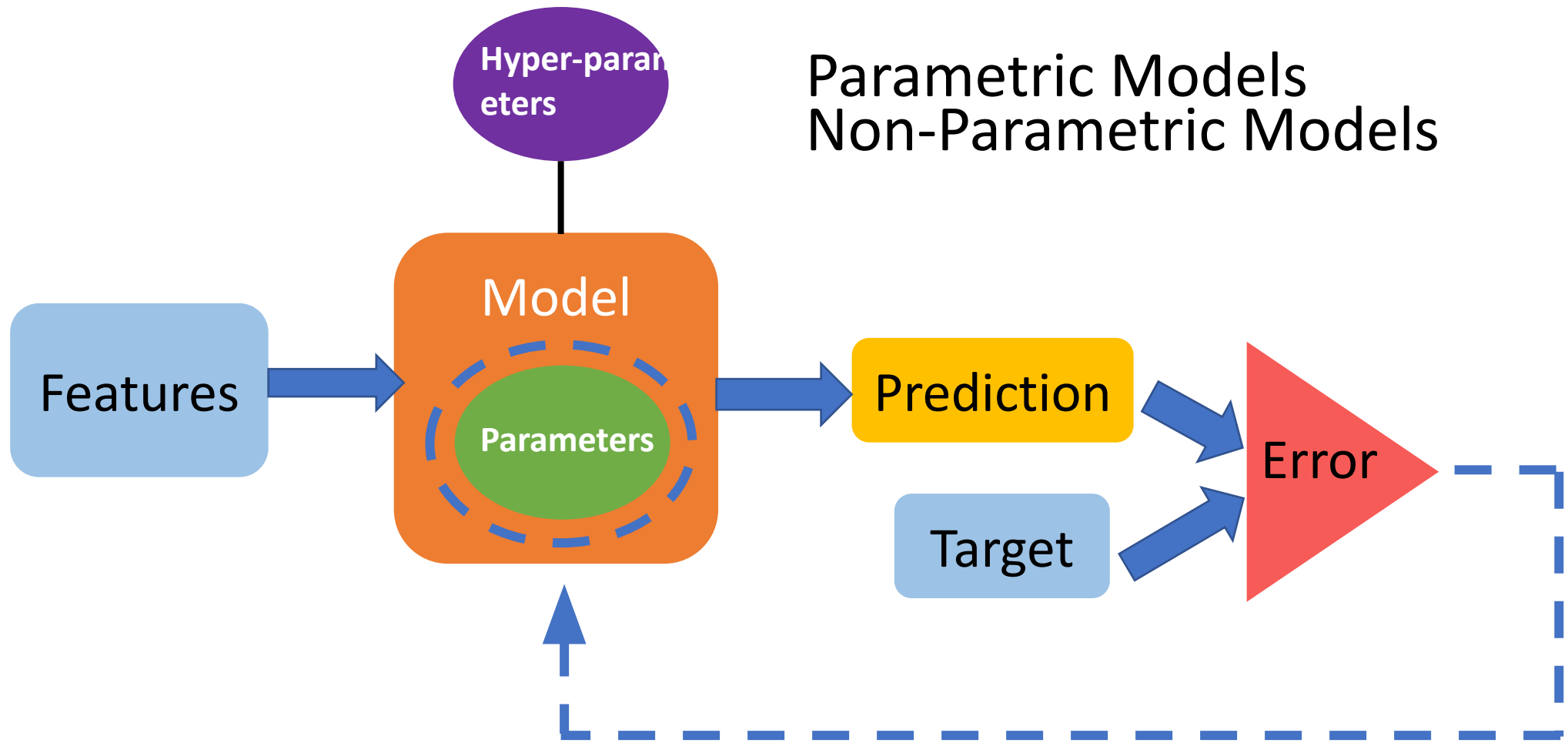


Non-parametric Models

Intro

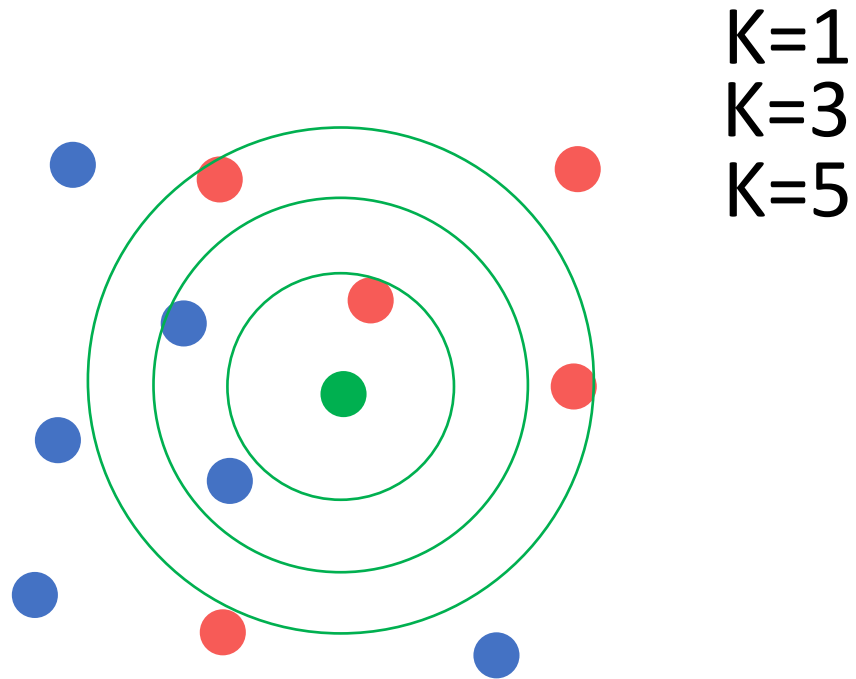
Non-parametric models



Non-parametric models

- K-Nearest Neighbors
- Decision Trees
- Support Vector Machine

K-Nearest Neighbors

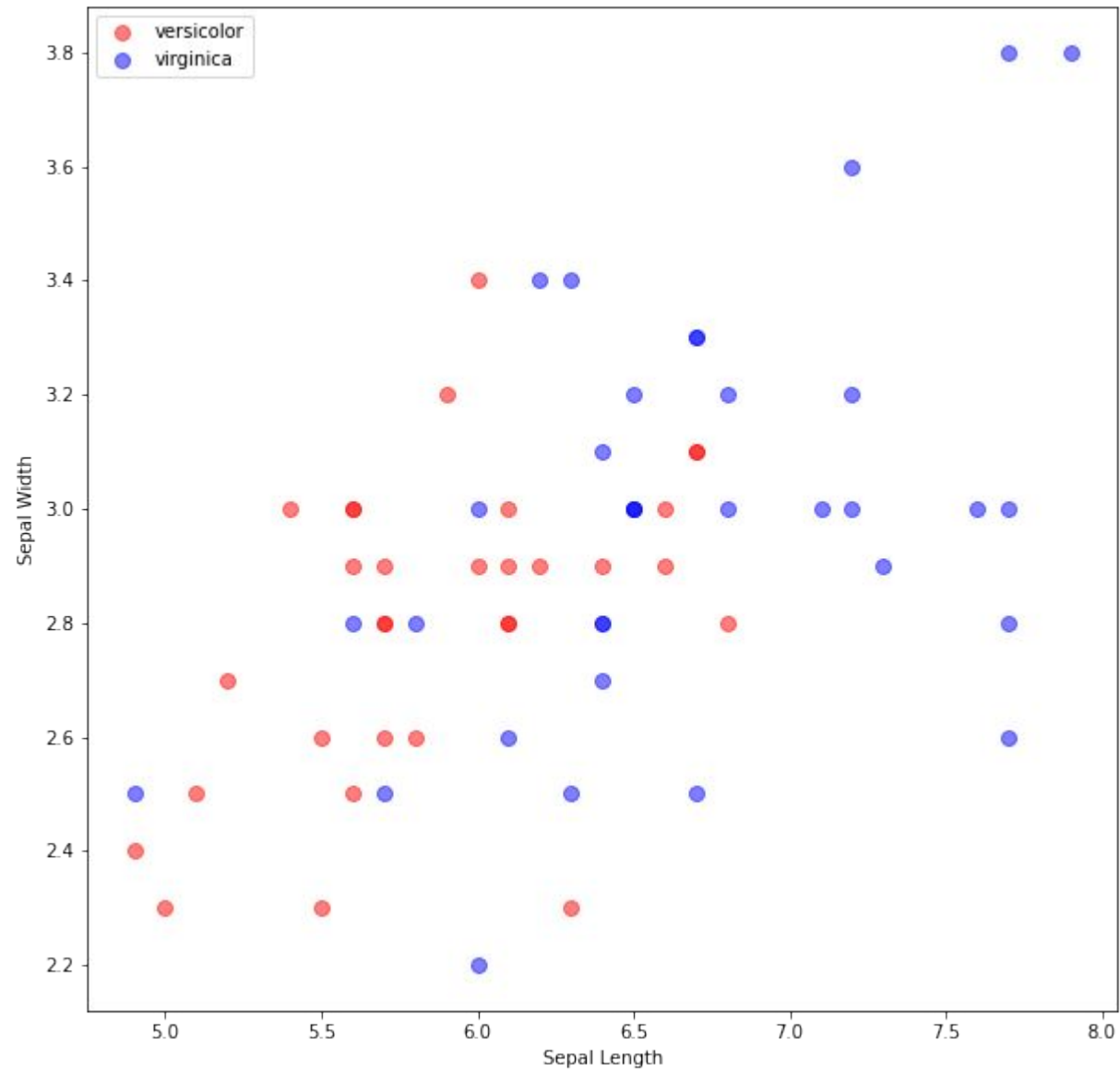


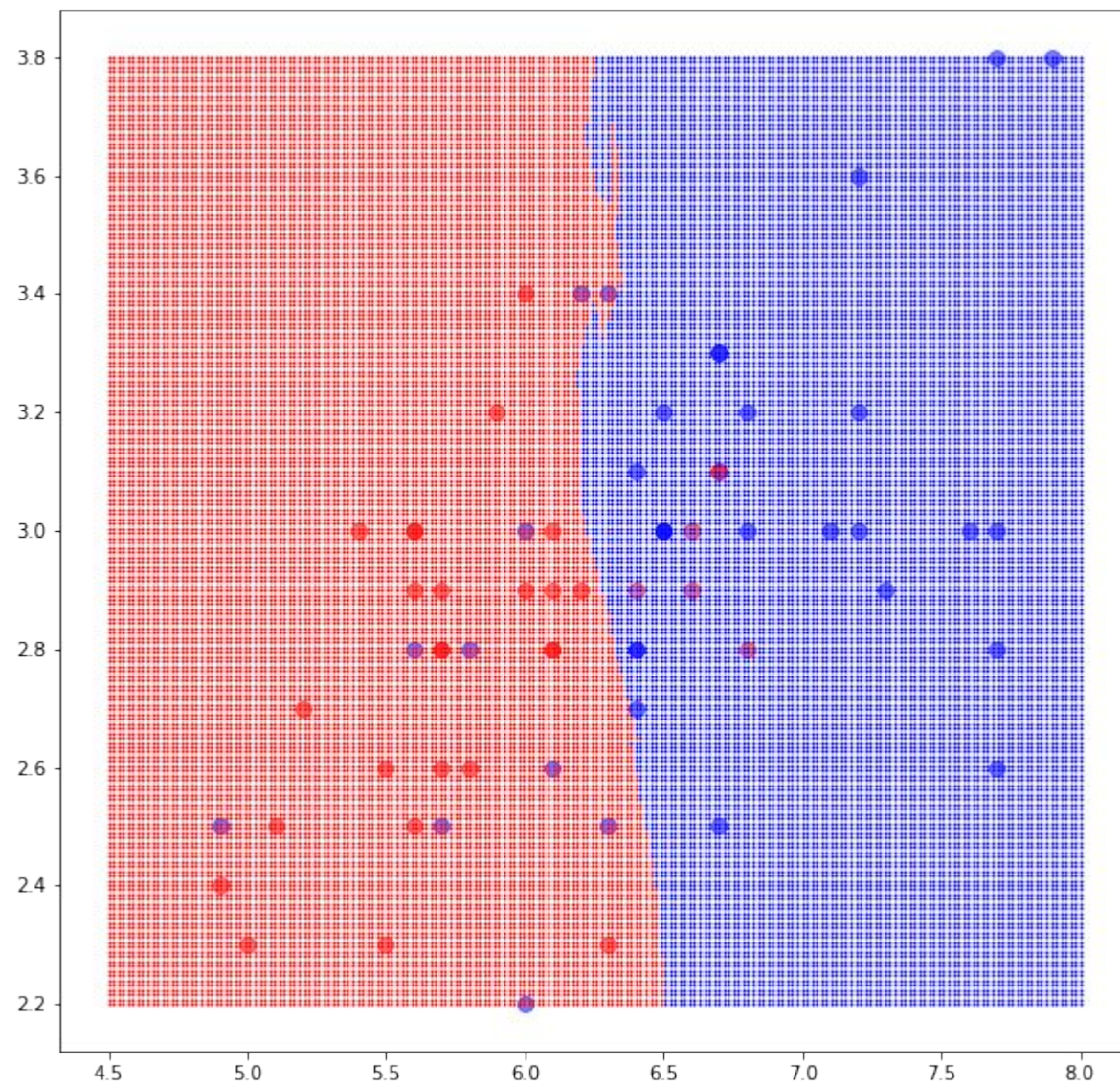
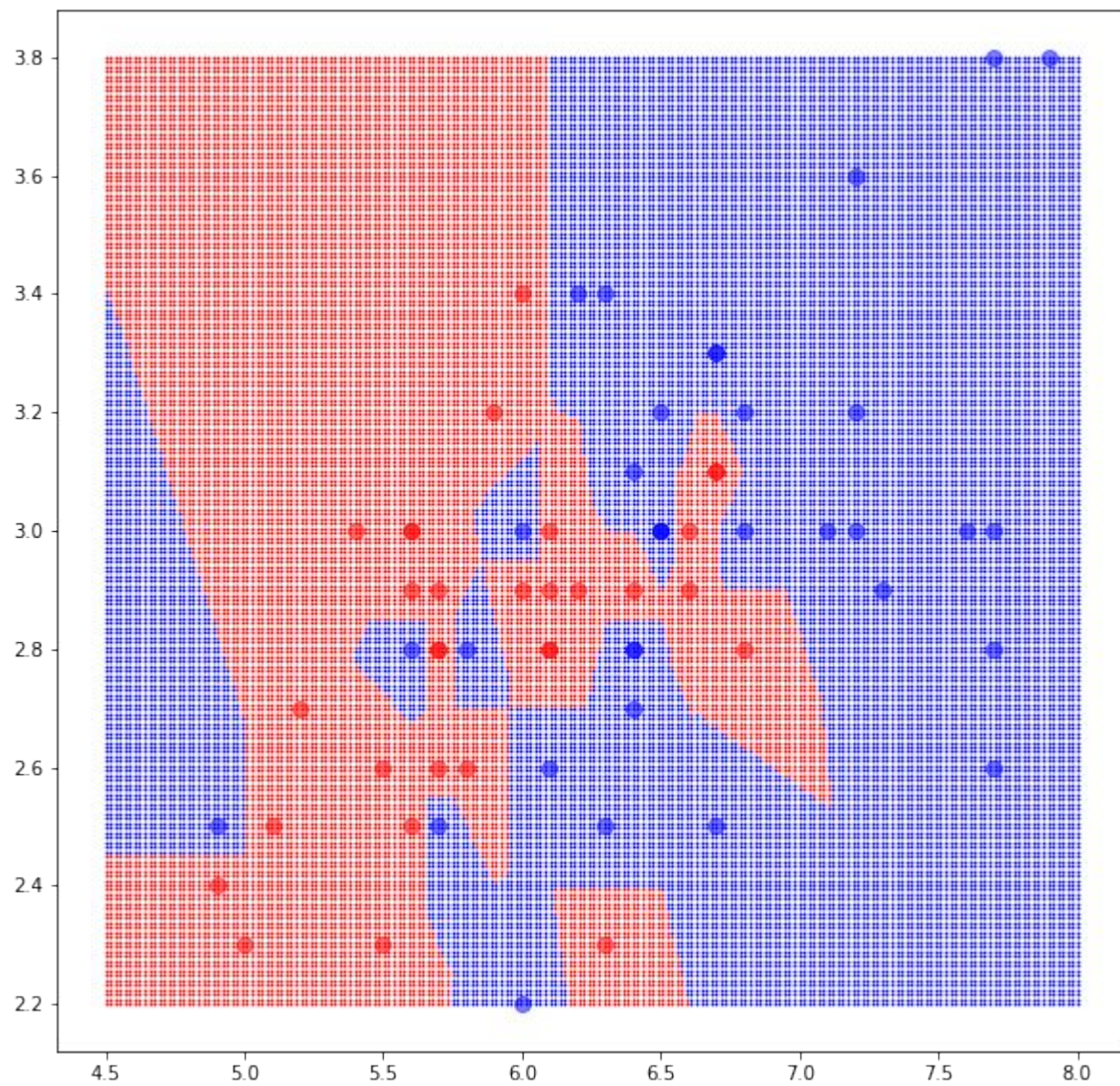
Manhattan distance

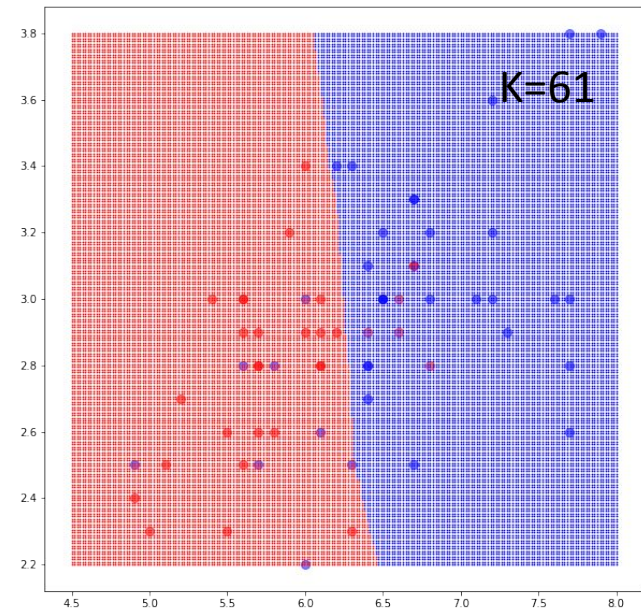
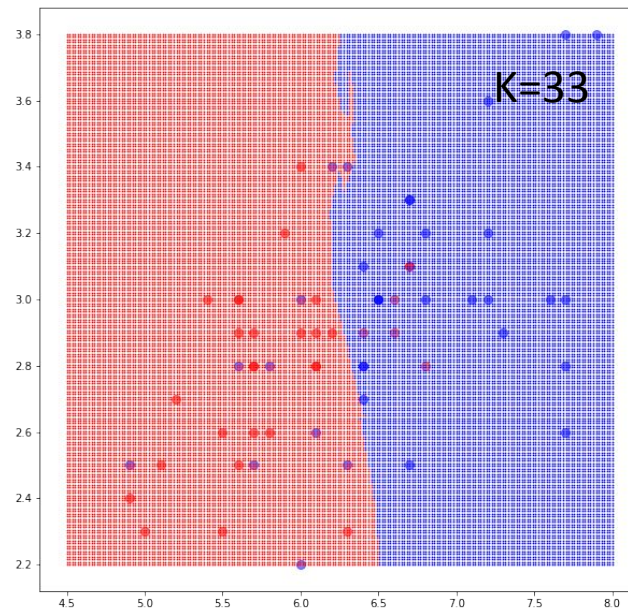
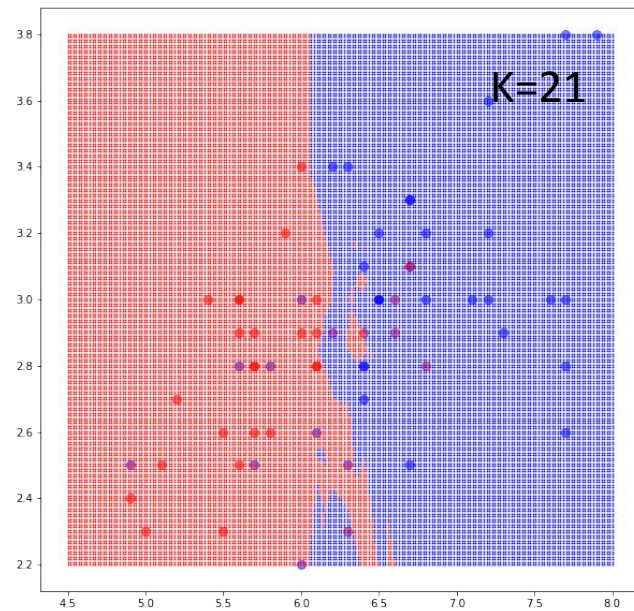
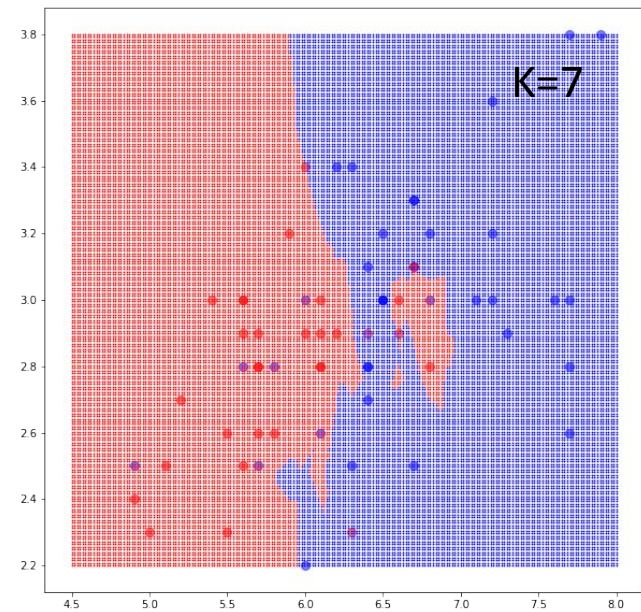
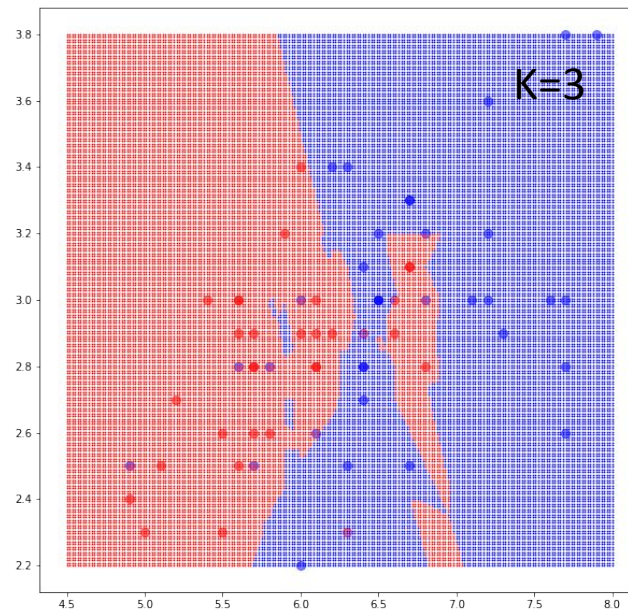
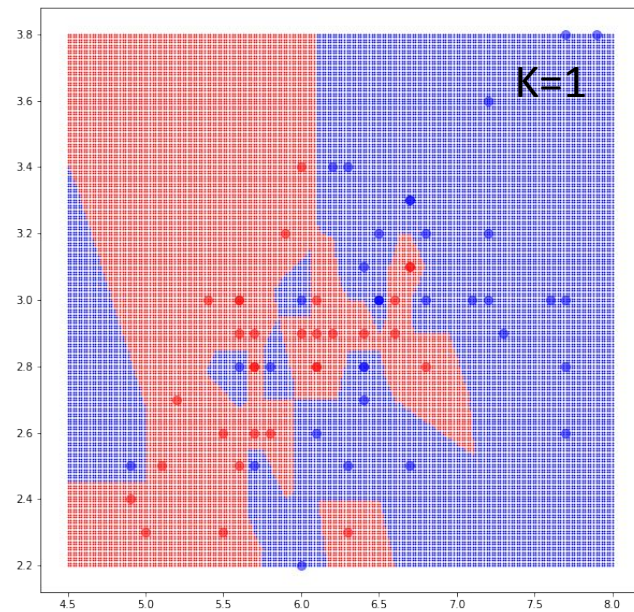
$$||\vec{X}_1 - \vec{X}_2||_1$$

Euclidean distance

$$||\vec{X}_1 - \vec{X}_2||_2$$







Bias and Variance of KNN

1. Which model has larger bias?

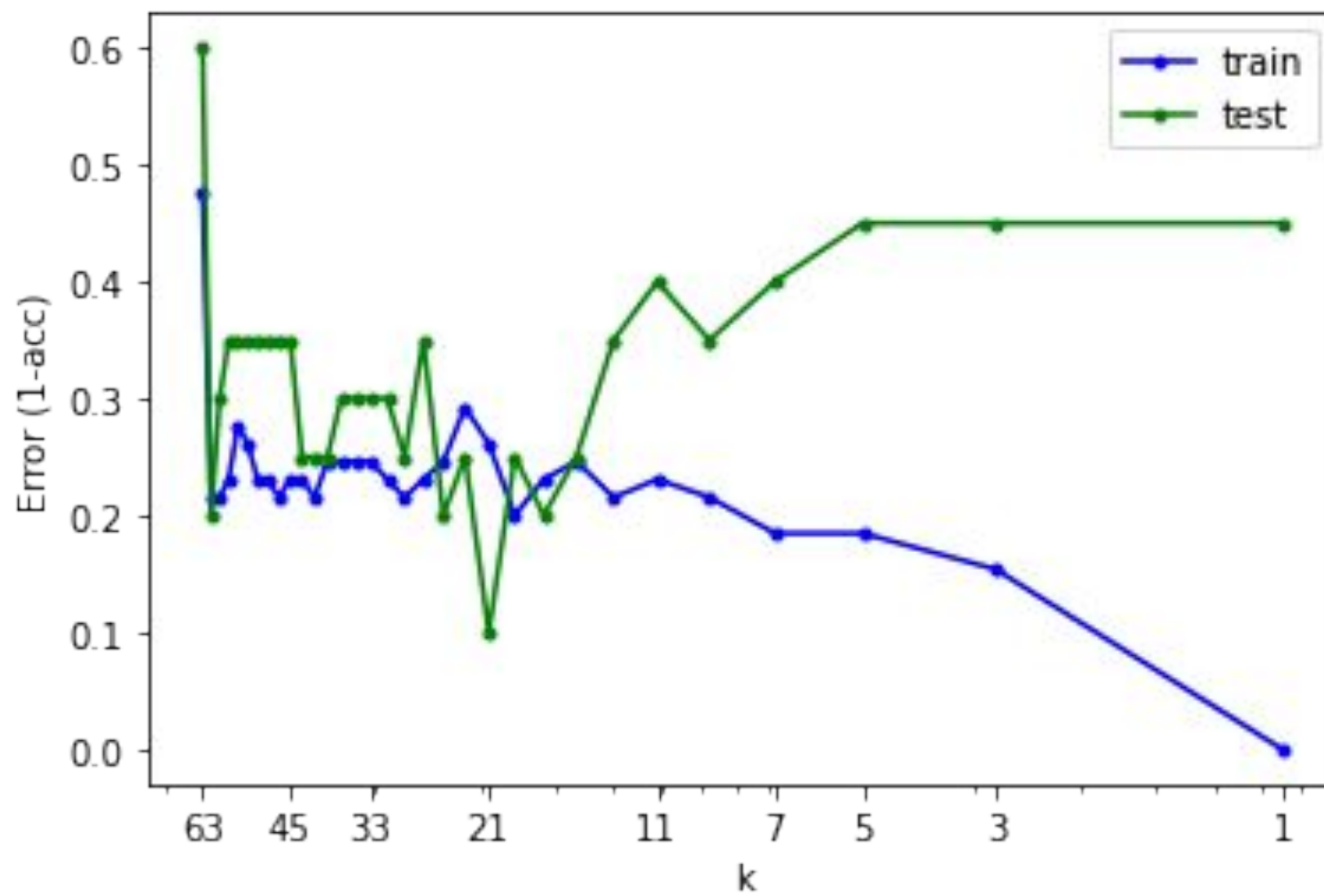
a. smaller k

b. larger k

2. Which model has larger variance?

a. smaller k

b. larger k



KNN properties

- Simple, memory-based algorithm
- Time complexity $O(nm)$

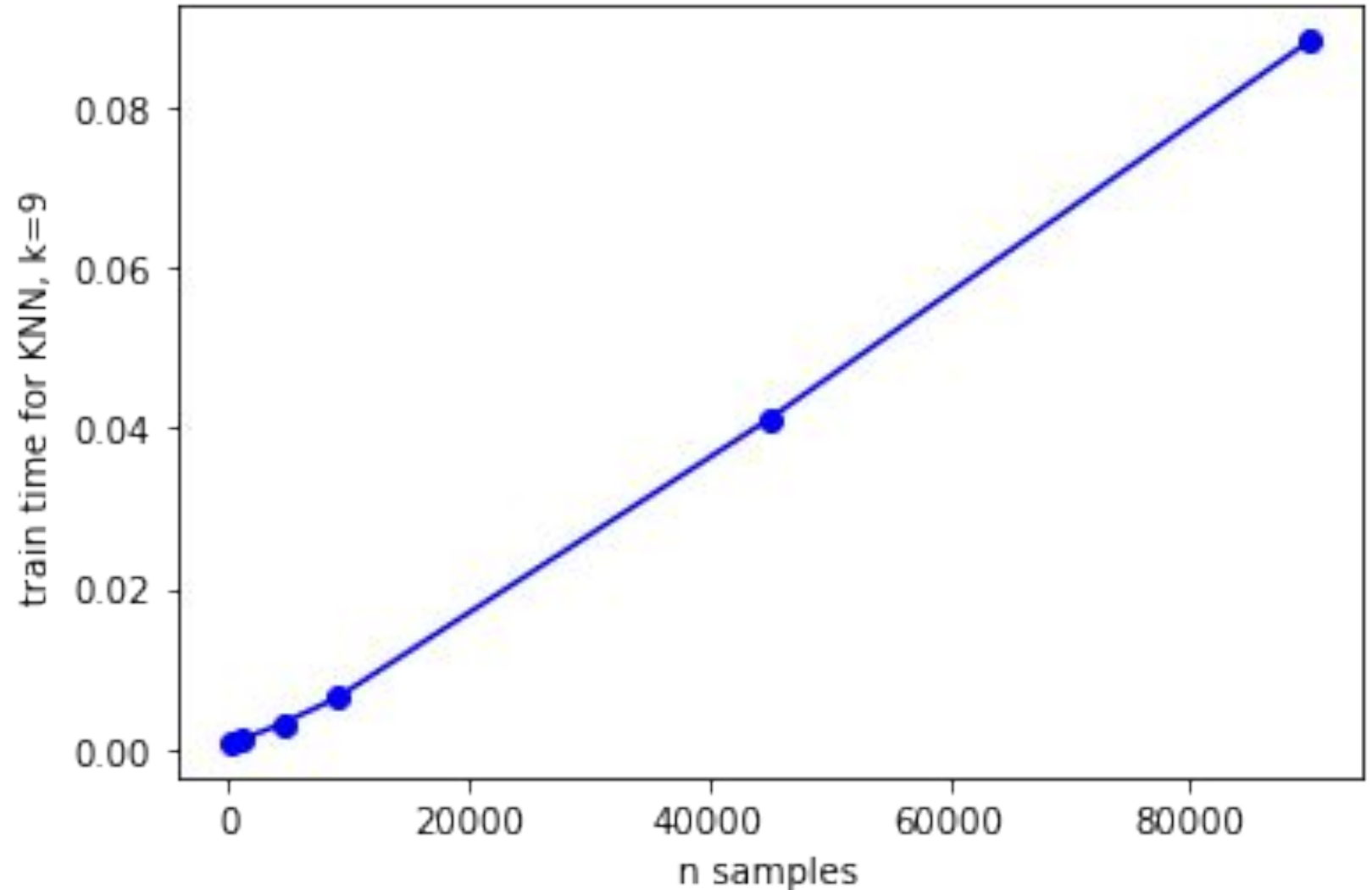
KNN properties

SensIT Vehicle dataset

90k+ samples

100 features

binary class

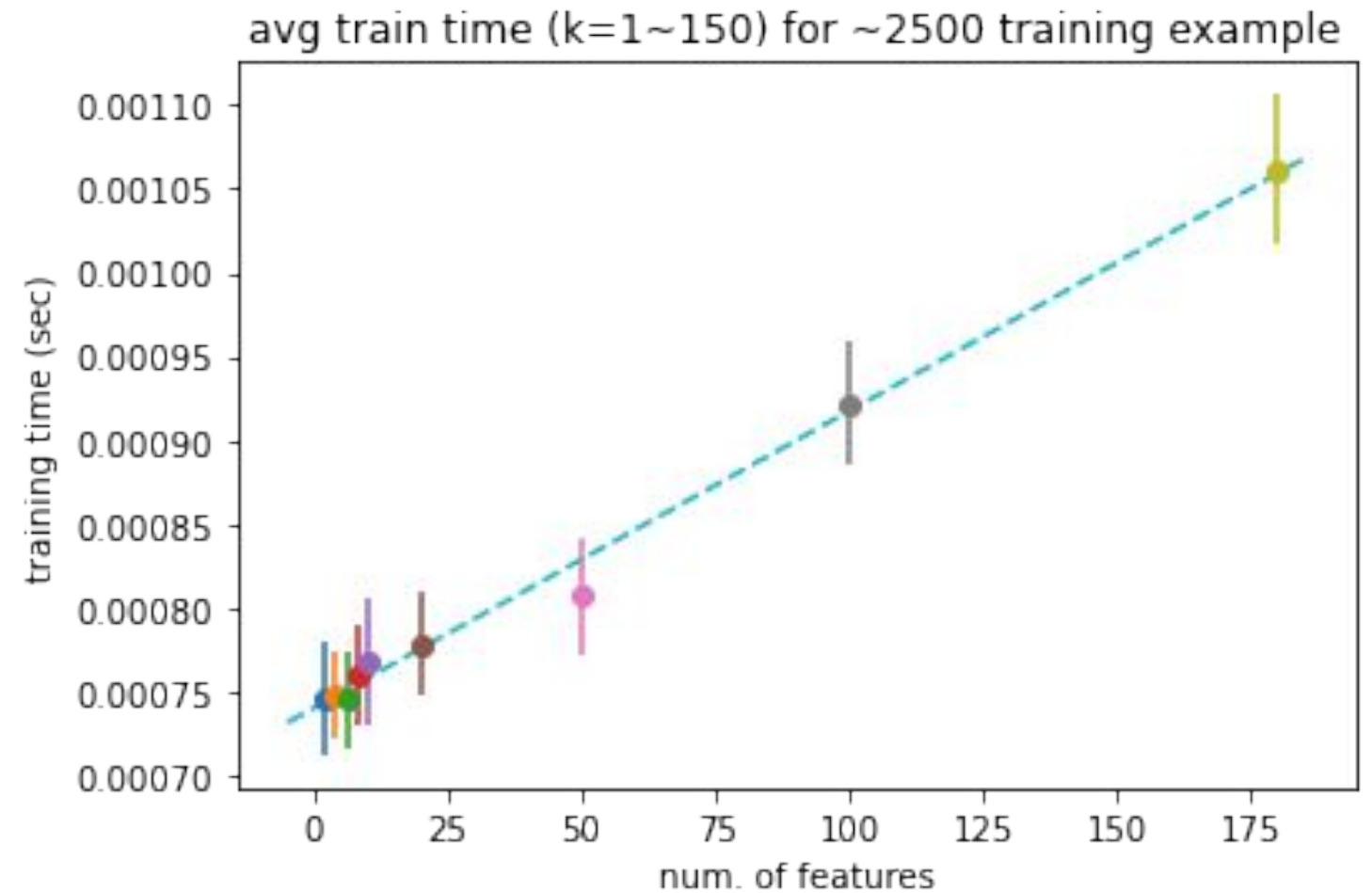


KNN properties- Example

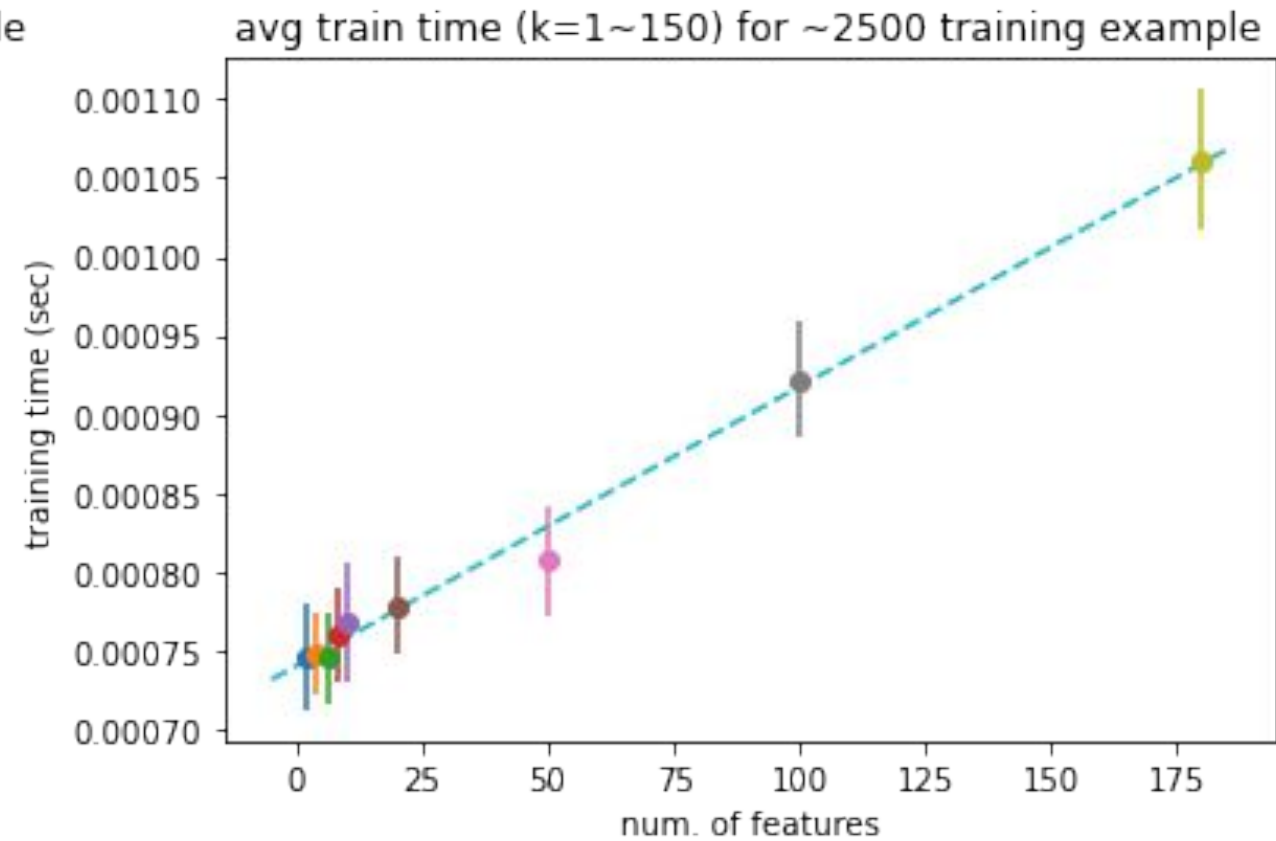
Splice-Junction Gene Sequences (DNA) dataset

180 binary features coding 60 ACGT nucleotide

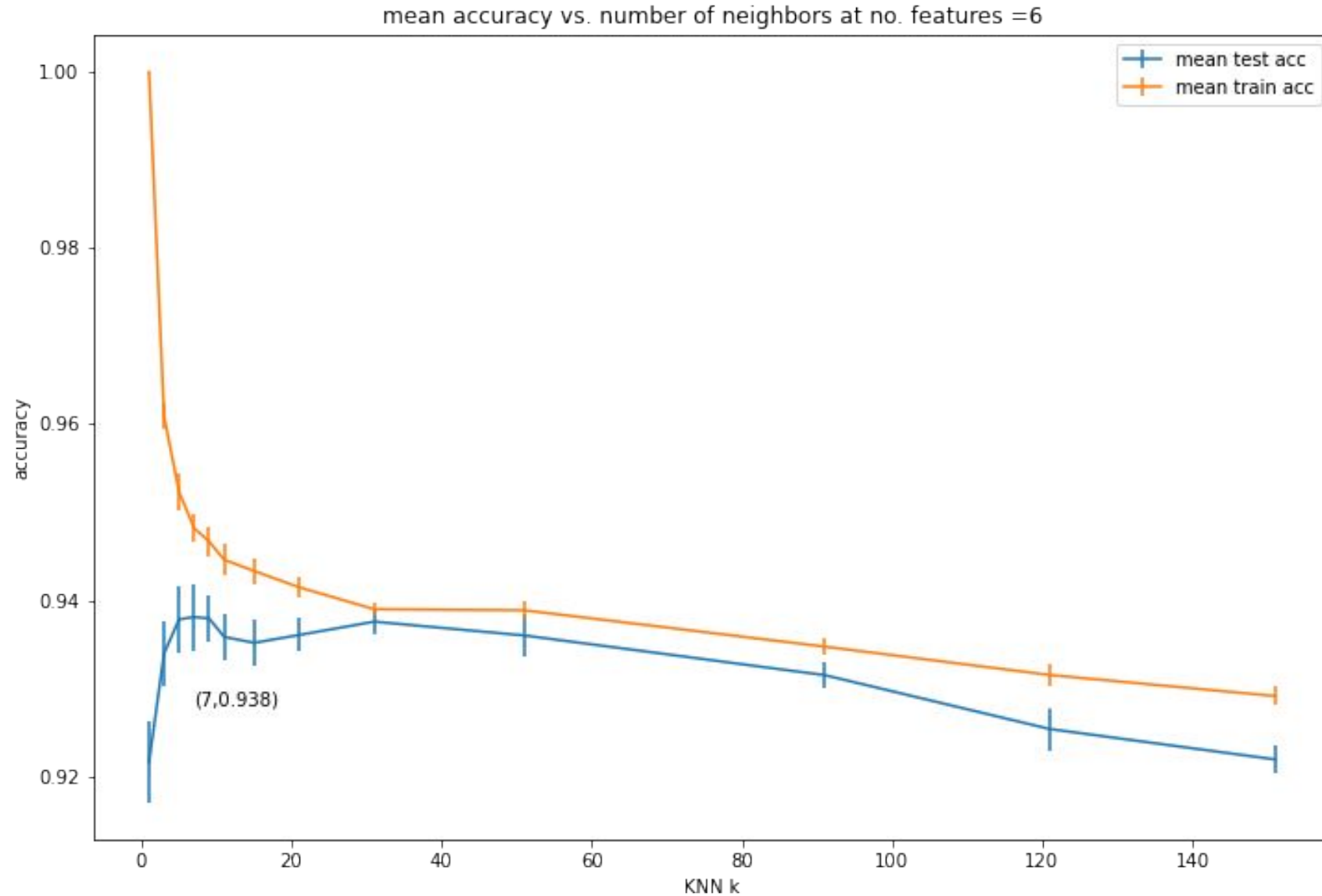
3 classes (exon boundaries)



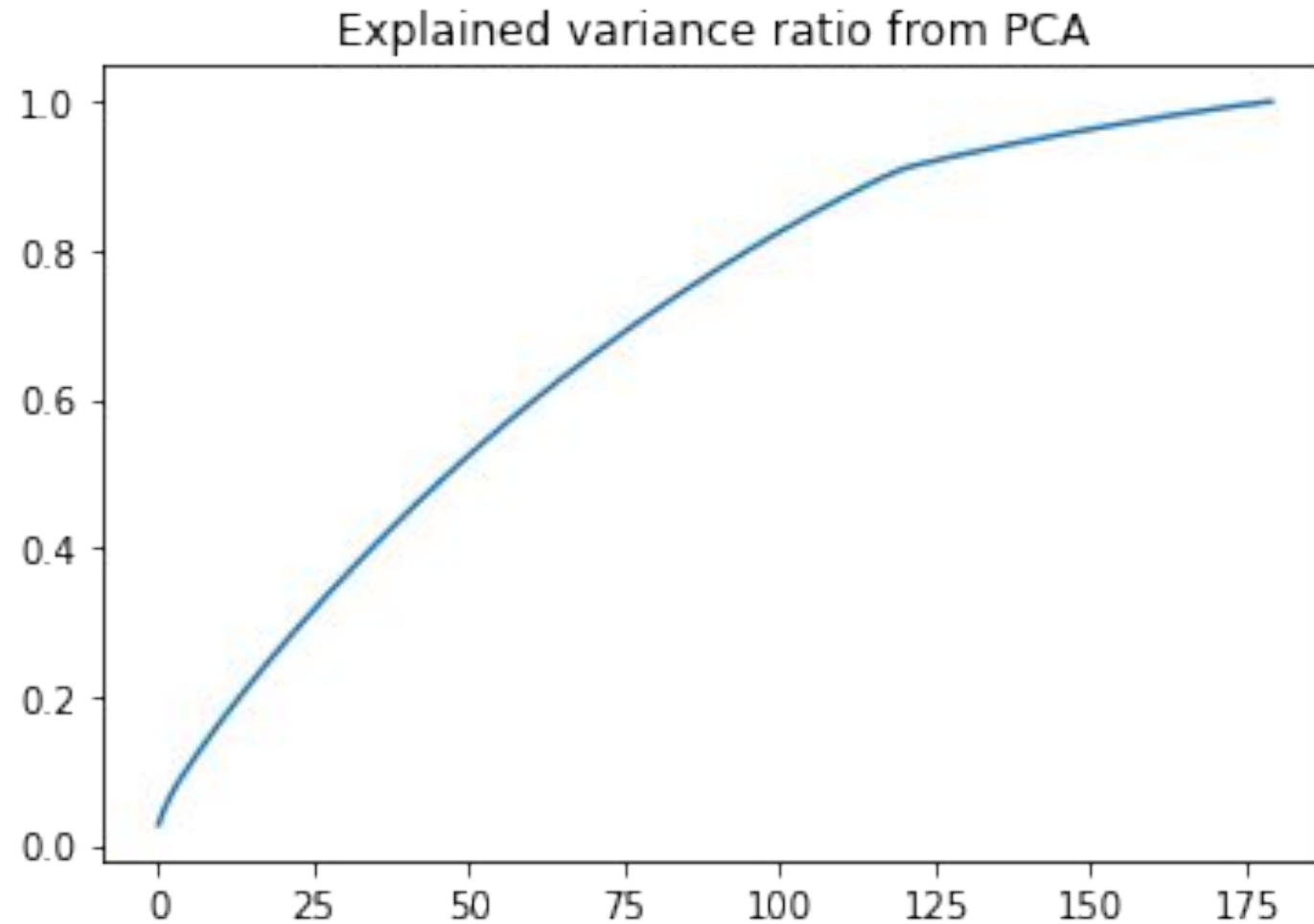
KNN properties



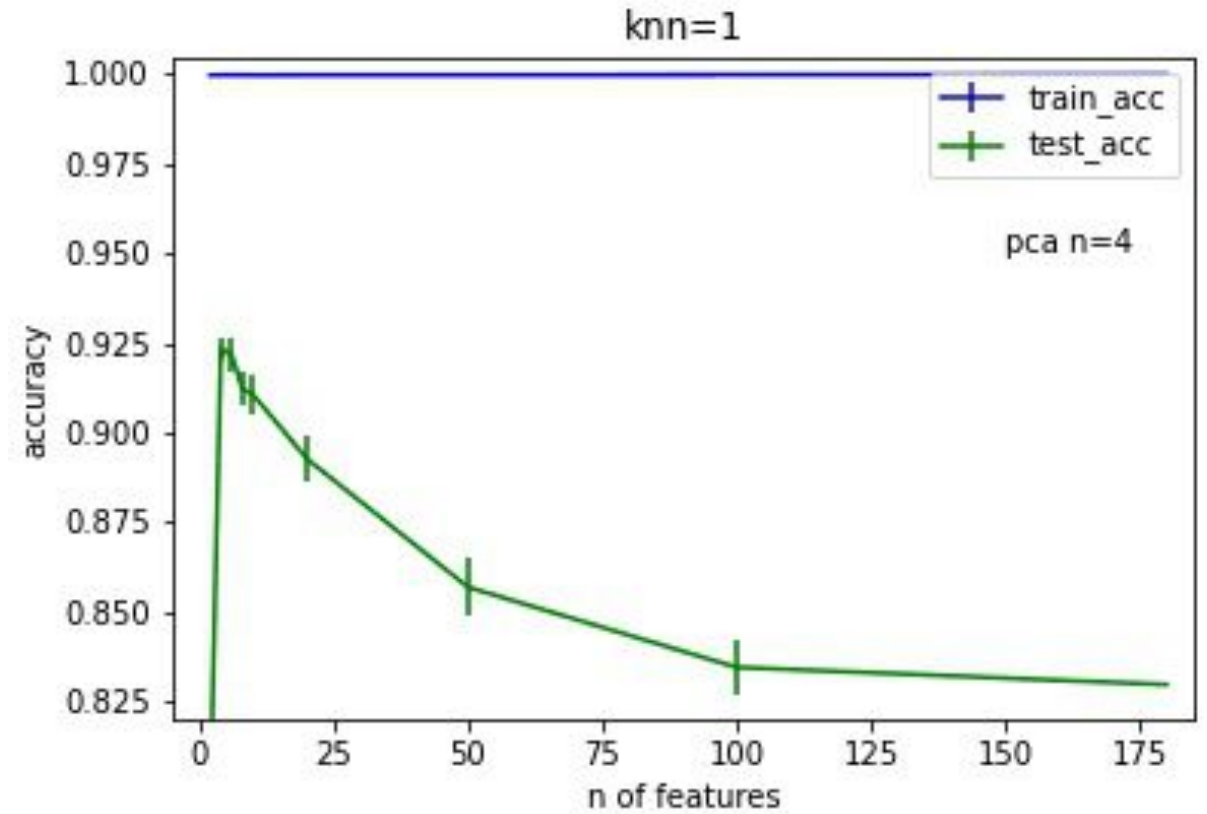
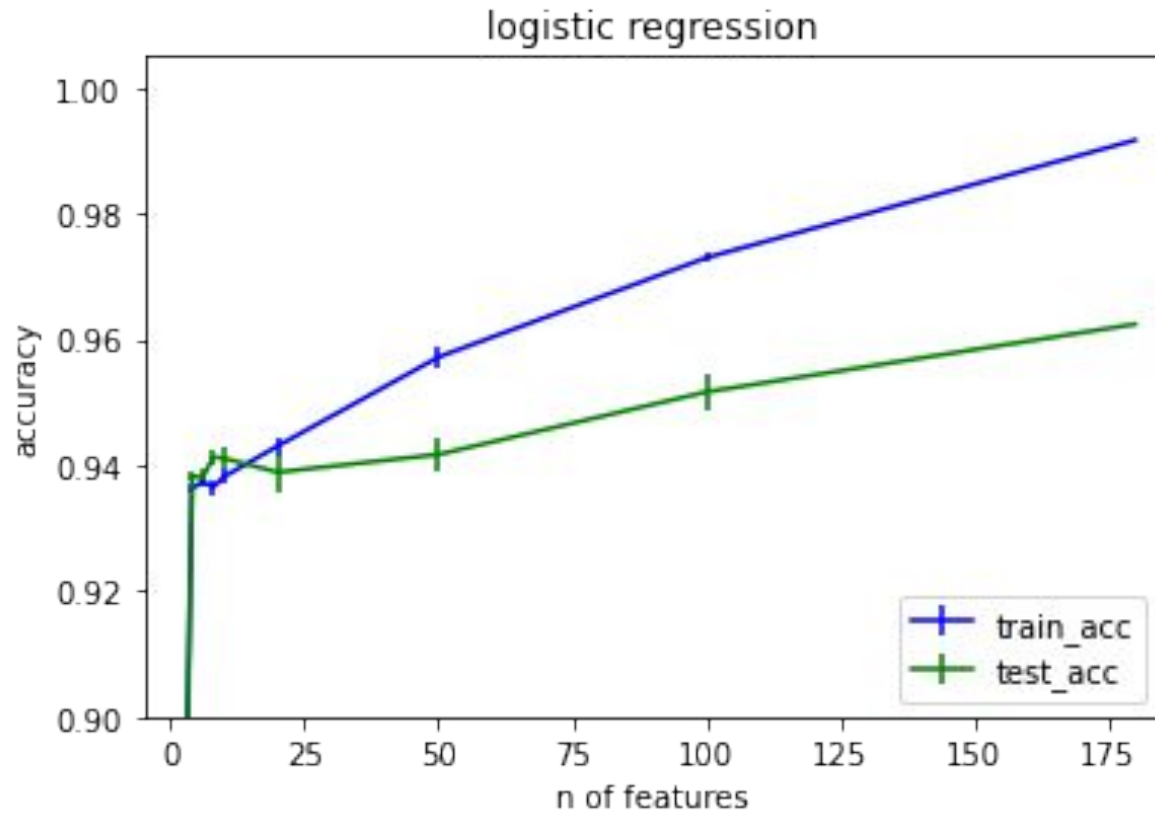
KNN properties



KNN Curse of dimensionality?



KNN Curse of dimensionality?



KNN Curse of dimensionality?

