K Nearest Neighbour

Anshu Pandey

Objectives

After completing this module, you should be able to understand:

- K Nearest Neighbour
- 1 Nearest Neighbour for Classification
- 1 nearest Neighbour for Regression
- Model Complexity
- Overfitting

K Nearest Neighbours

Conceptually one of the simplest Machine Learning algorithms.

Uses the proximity or similarity of observations to make predictions about them

Method:

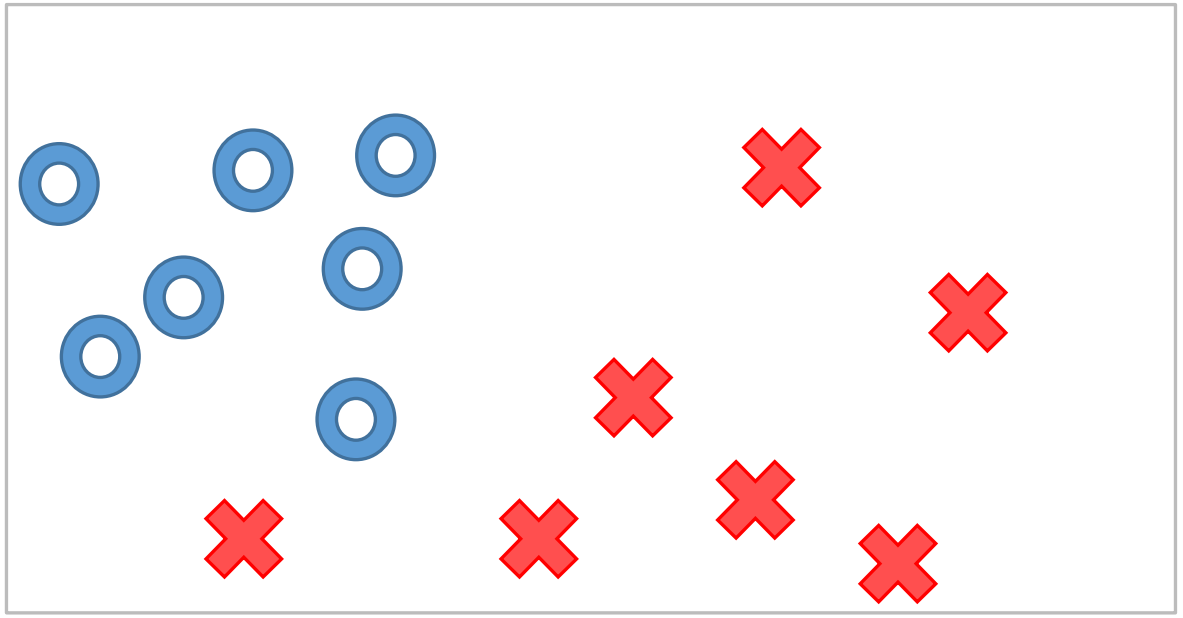
For the 1-Nearest Neighbour algorithm, find the closest labelled observation to the unlabelled observation and apply the same label.

For k Nearest Neighbour, find k closest labelled data points, take majority vote.

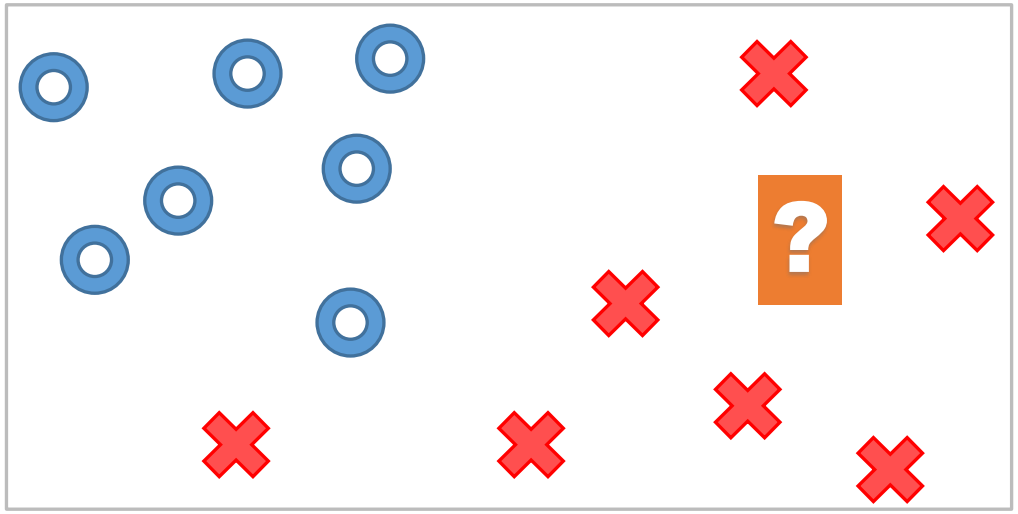
While it may seem very simple, it is often very effective!

It can be used for classification or regression

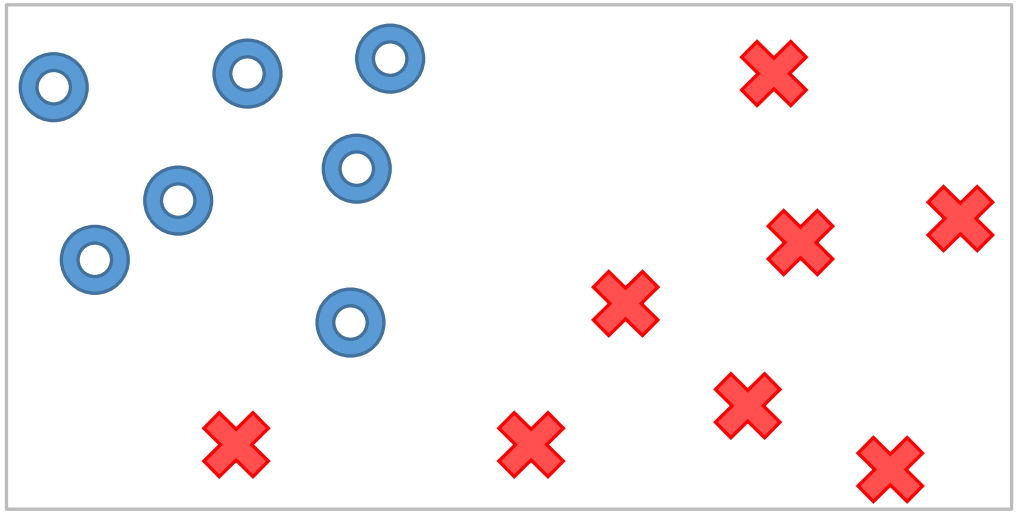
1 Nearest Neighbour predictions



1 Nearest Neighbour predictions

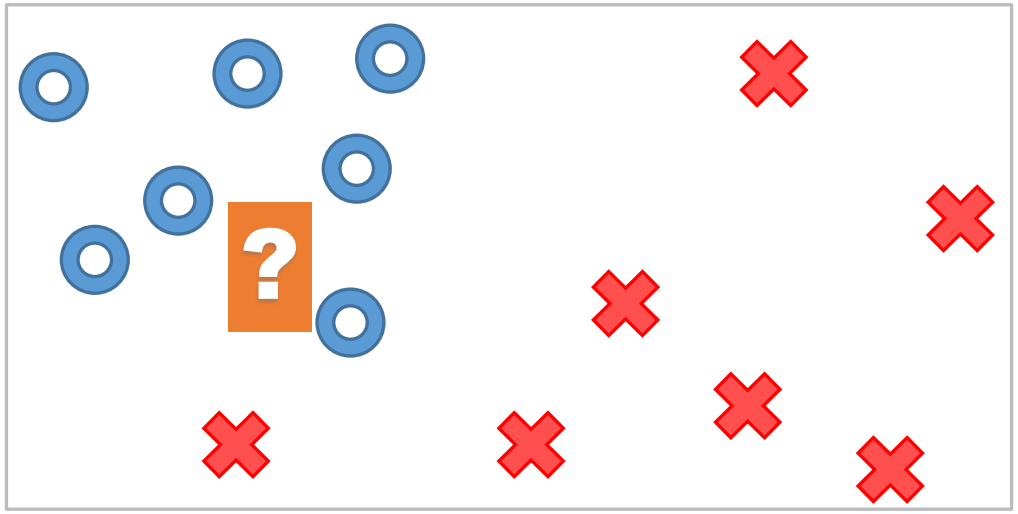


1 Nearest Neighbour predictions

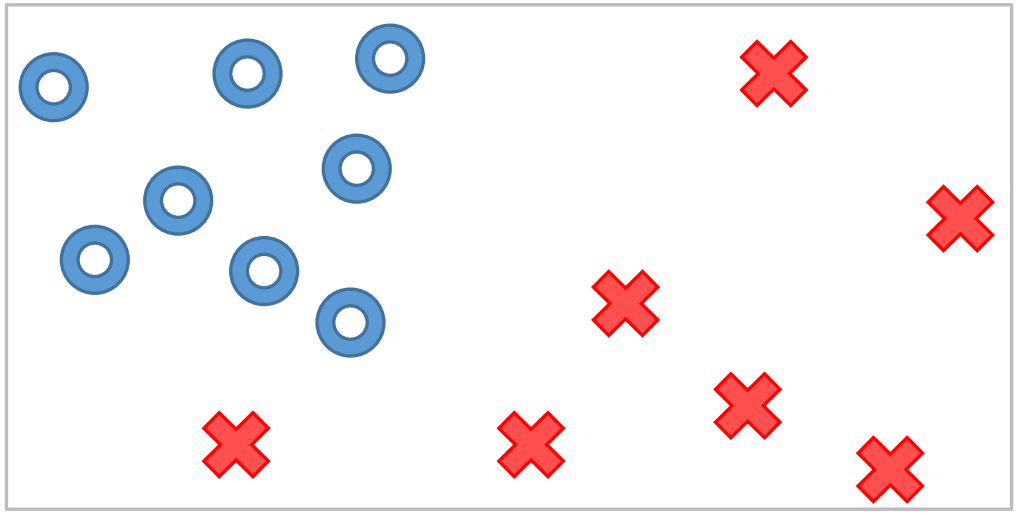


Anshu Pandey

1 Nearest Neighbour predictions

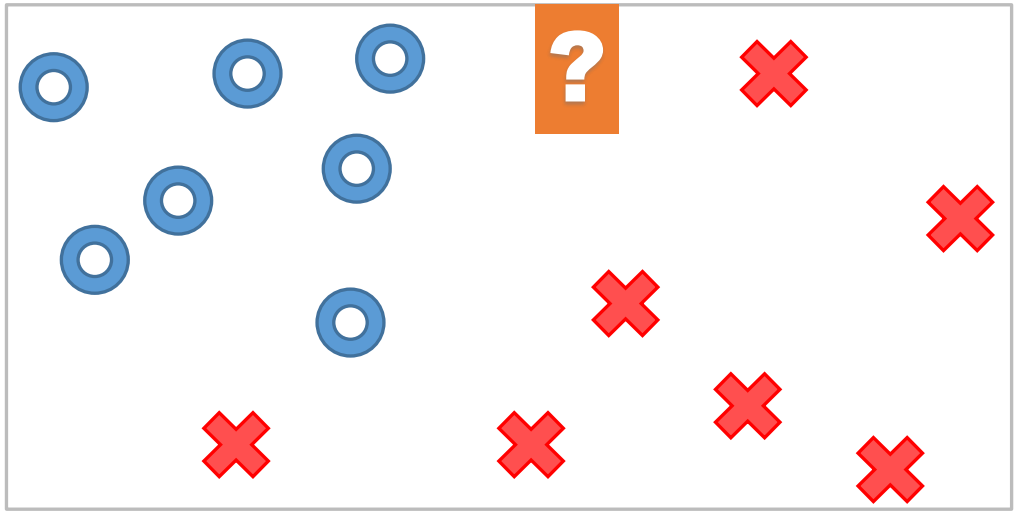


1 Nearest Neighbour predictions

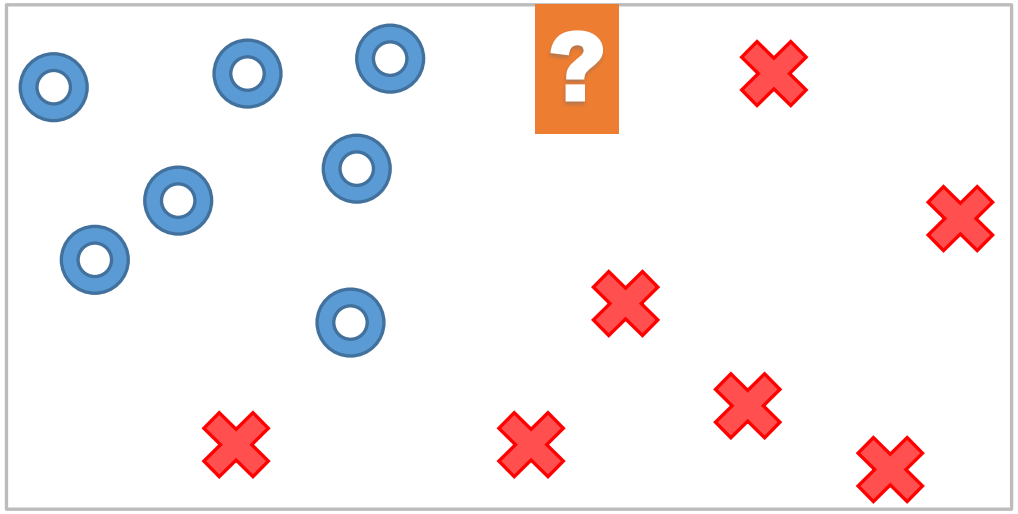


Anshu Pandey

1 Nearest Neighbour predictions



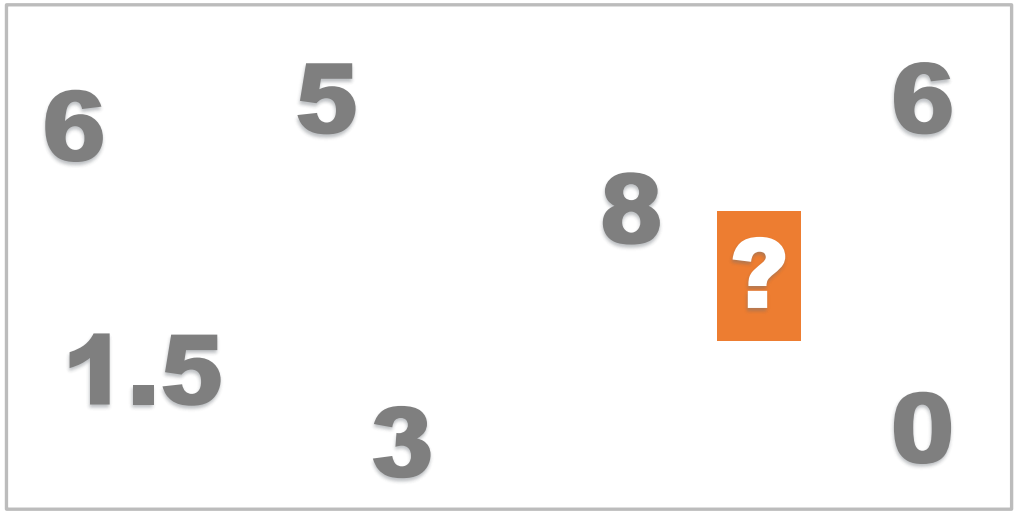
1 Nearest Neighbour predictions



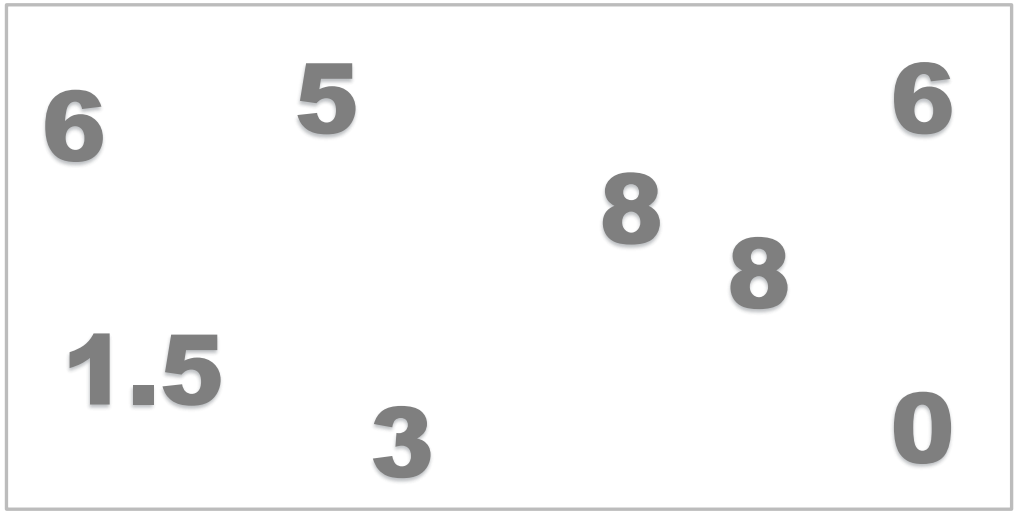
Here there is some ambiguity. We are equal distance from both classes.

In this case, for 1-NN we would just flip a coin to choose a class at random

Regression
1 Nearest Neighbour predictions



Regression
1 Nearest Neighbour predictions

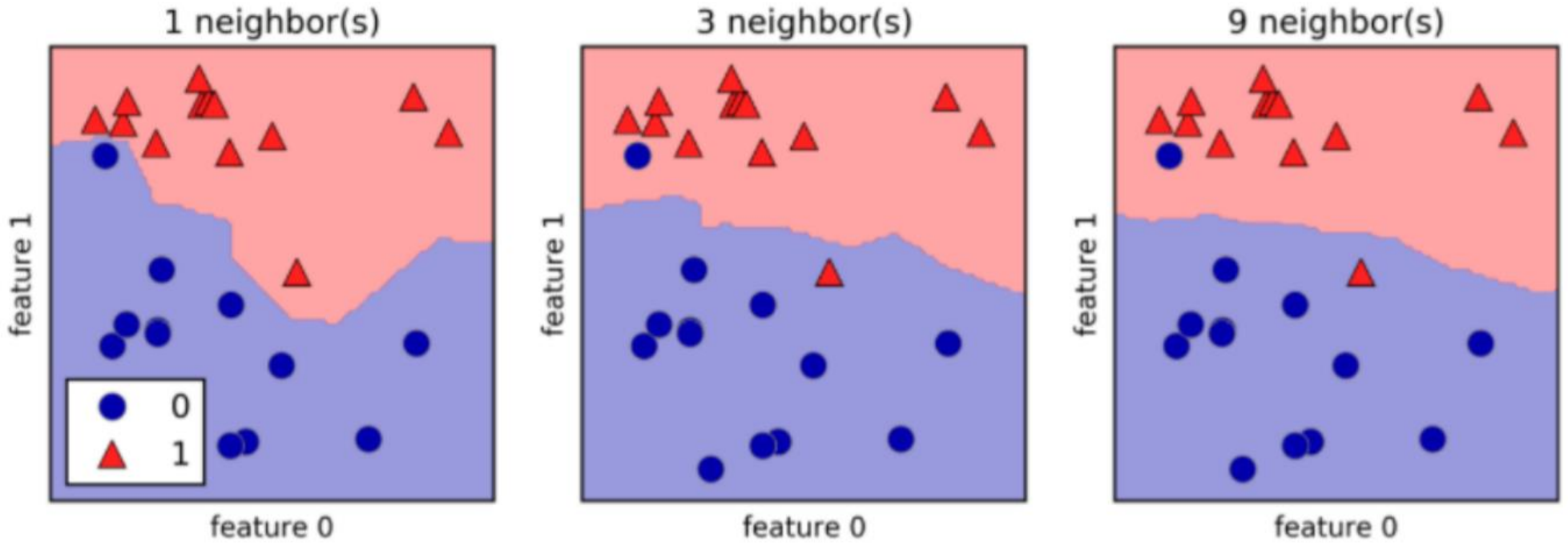


Model complexity

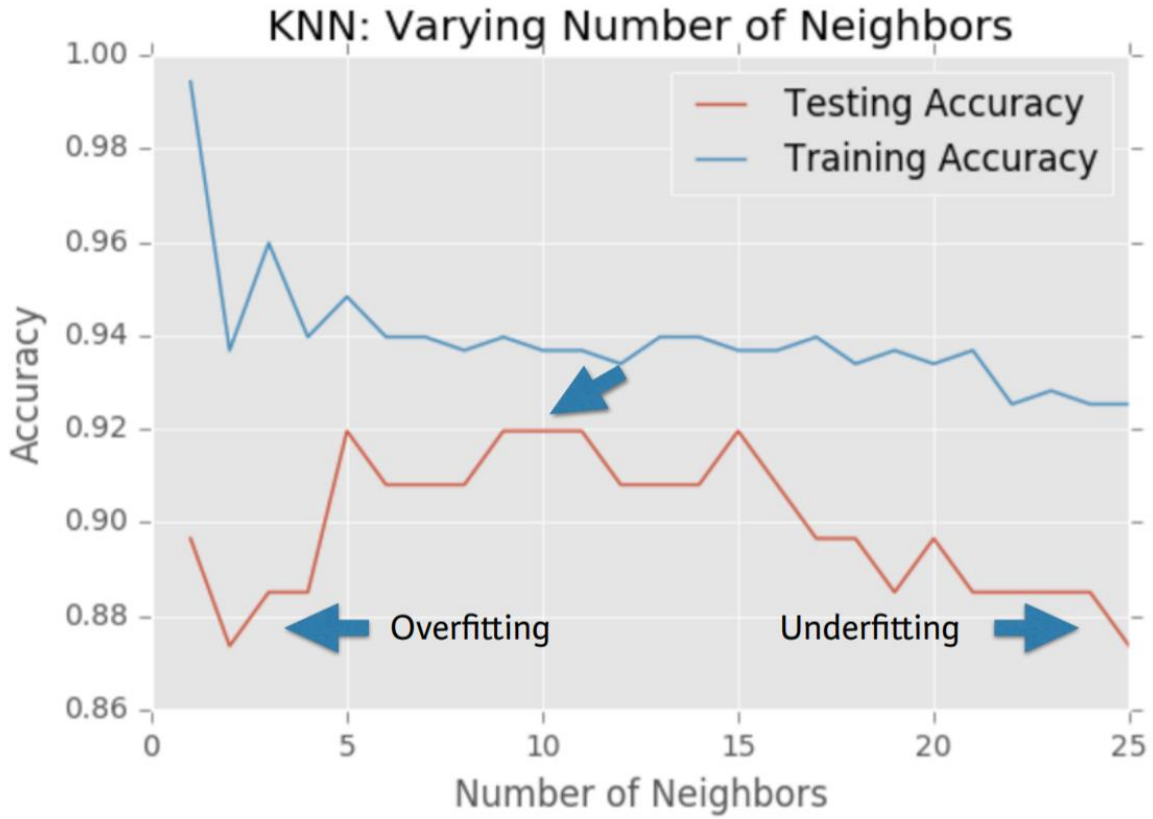
Larger k = smoother decision boundary = less complex model

Smaller k = more complex model = can lead to overfitting

Model complexity



Model complexity and Overfitting/Underfitting



Summary

This module covered the following topics:

K Nearest Neighbour

1 Nearest Neighbour for Classification

1 nearest Neighbour for Regression

Model Complexity

Overfitting

Anshu Pandey

Thank you

Anshu Pandey