

# K-Nearest Neighbors

Sauman Das



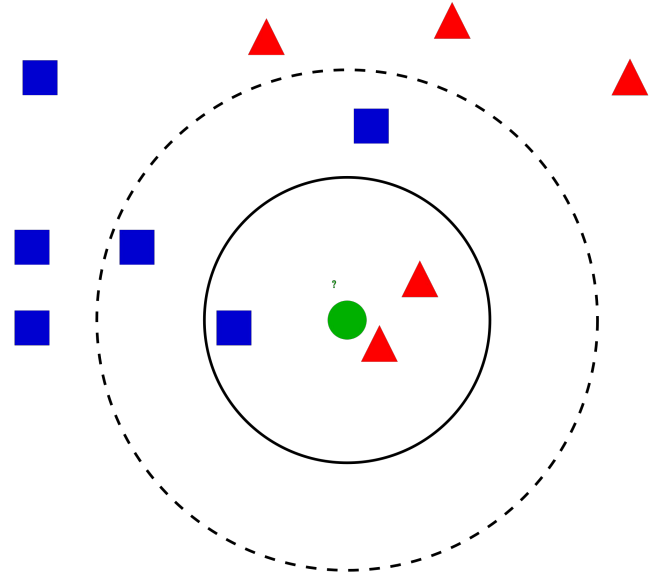
# Overview

- KNN Algorithm
- Pros and Cons
- Application



# Classification Algorithm

- Given test instance  $x$ 
  - Identify  $k$  closest points to  $x$  from the training set
  - Identify classes of the  $k$  closest points
  - Predicted class is the most common of the  $k$  classes



Did we create a  
model?



# Lazy Learner

- No actual “model” is created
- Use the entire training dataset for every test instance
  - How does this compare to decision tree (which is actually a model)?



# Pros

- Very simple model
  - Try implementing from scratch!
- No training time
  - Enables us to adaptively increase training data

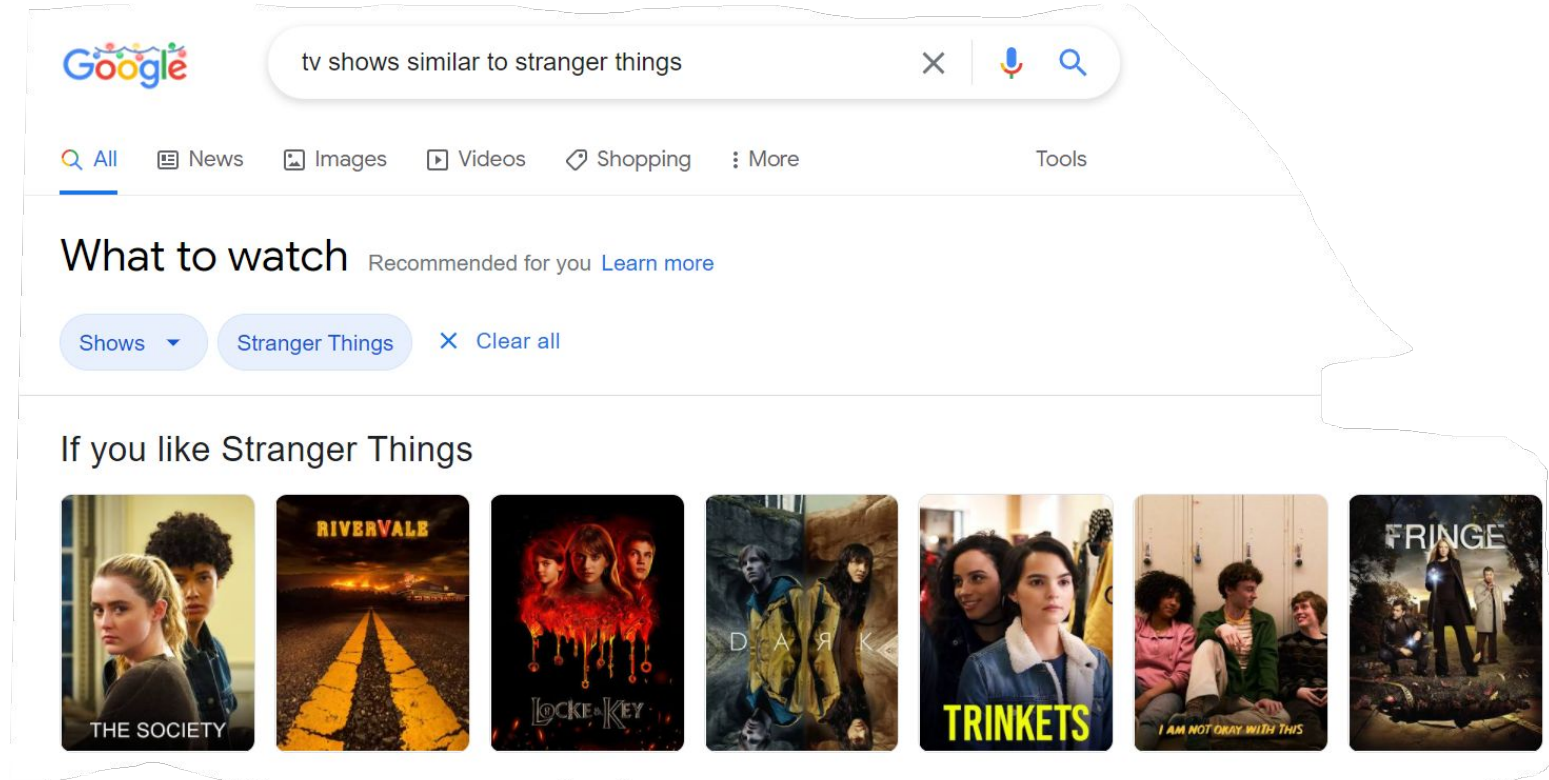


# Cons

- Large amount of memory
  - Stores entire training dataset!
- Runtime for each test case is large
- Simple algorithm could lead to poor results



# Recommender Systems



The image shows a Google search interface. The search bar contains the text "tv shows similar to stranger things". Below the search bar, there are navigation links for "All", "News", "Images", "Videos", "Shopping", and "More", along with a "Tools" link. The main content area is titled "What to watch" and includes a sub-header "Recommended for you" with a "Learn more" link. Below this, there are filter buttons for "Shows", "Stranger Things", and "Clear all". The main recommendation section is titled "If you like Stranger Things" and displays a row of seven show posters: "THE SOCIETY", "RIVERVALE", "LOCKE & KEY", "D.E.A.R.", "TRINKETS", "I AM NOT OKAY WITH THIS", and "FRINGE".

Google

tv shows similar to stranger things

All News Images Videos Shopping More Tools

What to watch Recommended for you [Learn more](#)

Shows Stranger Things Clear all

If you like Stranger Things

THE SOCIETY RIVERVALE LOCKE & KEY D.E.A.R. TRINKETS I AM NOT OKAY WITH THIS FRINGE





# Recommender System

- Each movie has associated features
- If you want 3 of the most similar movies, run KNN with  $k=3$
- Output is the closest neighbors
  - Not an actual class value in this case

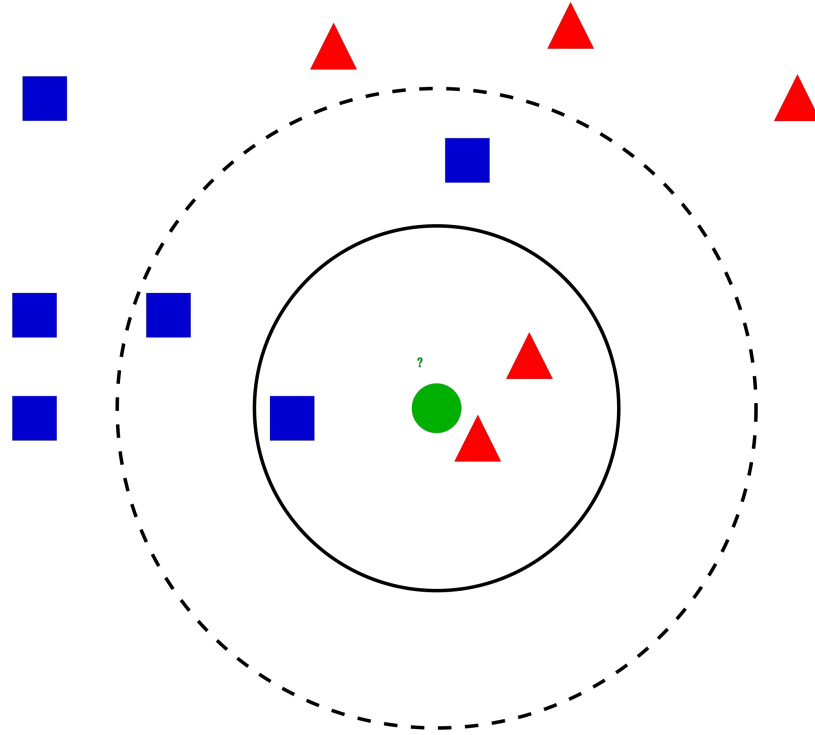


# Recommender System

- KNN is useful in this scenario
- As new movies are added frequently to the database, we don't retrain the model
- Works as long as there is a database of movies (training set)

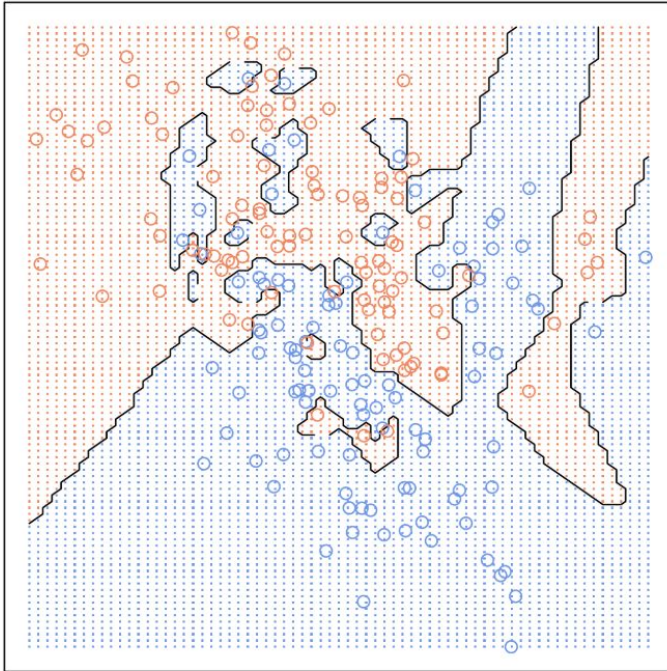


# Overfitting



# Increasing $k$

1-nearest neighbours



20-nearest neighbours

