MATH 156 Final Project

Predicting videogame sales with various models

Group 5

University of California, Los Angeles

July 27, 2020

Table of Contents

Table of Contents

K Nearest Neighbor Regression

Goal: given $x \in \mathbb{R}^d$, predict sales.

- Find k nearest data points to x.
- Compute the predicted sales based on these k points.

```
from sklearn.neighbors import KNeighborsRegressor
model = KNeighborsRegressor(n_neighbors=k).fit(X_train,
      Y_train)
res = model.predict(X_test, Y_test)
```

K Nearest Neighbor Regression

Goal: given $x \in \mathbb{R}^d$, predict sales.

- Find k nearest data points to x.
- Compute the predicted sales based on these k points.

```
from sklearn.neighbors import KNeighborsRegressor
model = KNeighborsRegressor(n_neighbors=k).fit(X_train,
      Y_train)
res = model.predict(X_test, Y_test)
```

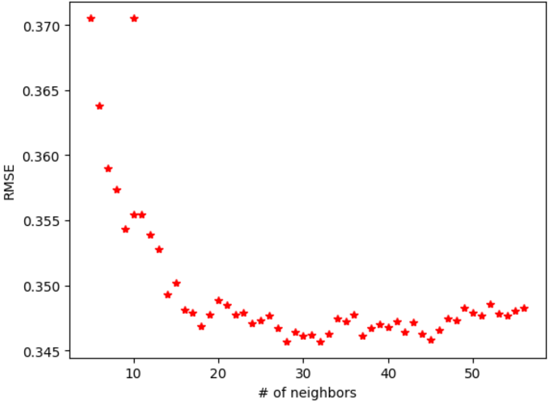
Questions we should think about:

- How to determine k?
- How to find the nearest points efficiently?
- How to predict the sales based on the points?

Cross Validation

How to determine k?

- Divide the training dataset into two parts (actual training and cross validation).
- Find the optimal k with the cross validation set.



KD-Tree

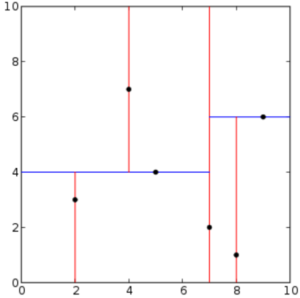
How to find the nearest points efficiently given that the training size is n and the test size k?

- Naive approach
  - Compare with all data points in the training set
  - Time Complexity: $\mathcal{O}(kn)$

KD-Tree

How to find the nearest points efficiently given that the training size is n and the test size k?

- Naive approach
  - Compare with all data points in the training set
  - Time Complexity: $\mathcal{O}(kn)$
- KD-Tree
  - Construct a KD-Tree and search
  - Time Complexity: $\mathcal{O}(k \log n)$.



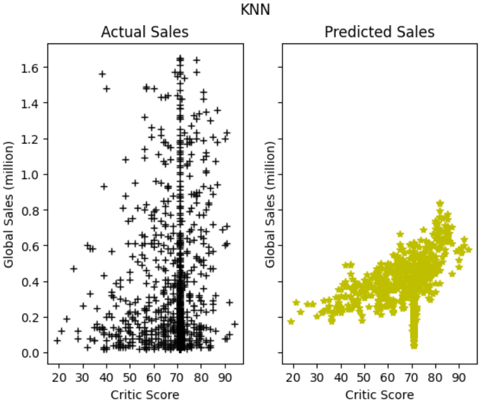
Sales Prediction

How to predict sales based on nearest points?

- Mean
- Median
- Linear Regression

Results

- Pros
  - No assumptions about the data
- Cons
  - Localized data when k increases
  - Memory inefficient and slow


KNN
Actual Sales
Predicted Sales
Global Sales (million)
Critic Score