IB Seminar – KNN Project

House Prediction

In housing price prediction, literally being a “near neighbor” is actually a good indicator of being similar in price. k-NN is useful in domains where physical proximity matters.

How to use k-NN to predict housing prices:

1. **Store the training data,** 
   1. matrix X features:
      1. zip code,
      2. town,
      3. # of bedrooms,
      4. # of bathrooms,
      5. square feet,
   2. matrix Y of corresponding sale prices.
2. **Sort the houses in your training data set by** **similarity** to the house in question, based on the features in X.
3. **Take the mean of the *k* closest houses.** That is your guess at the sale price (i.e. ŷ)

**Choosing k: tuning hyperparameters with cross-validation**

To decide which value of *k*to use, you can test different k-NN models using different values of *k* with **cross-validation:**

1. Split your training data into segments, and train your model on all but one of the segments; use the held-out segment as the “test” data.
2. See how your model performs by comparing your model’s predictions (ŷ) to the actual values of the test data (y).
3. Pick whichever yields the lowest error, on average, across all iterations.