Lecture 34
Enhancing Nearest Neighbor
Announcements

● Project 3 out, Part 1 due 5/7, Part 2 due 5/14.
● Labs the week of 5/3 and 5/10 will be dedicated to Project 3.
● Final May 22, 1:30PM-4PM, in Baker Lab 200
  ○ If you will be getting second vaccine shot May 20-22, let me know by May 7.
  ○ If 1:30PM-4PM Eastern is not in 8AM-10:30PM in your local time zone, let me know by May 7.
Announcements

Guest speakers Friday 5/7, Monday 5/10, Wednesday 5/12

● Yes, participation in these will still count (so please show up) and questions about the materials covered will be fair game for the final.

● Friday: Cornell COVID modeling team

● Monday: Professor Karen Levy, Information Science

● Wednesday: Dr. Ehi Nosakhare, Microsoft
Classification

- Our study of regression:
  - One quantitative variable (x)
  - Predicts another quantitative variable (y)

- Now, classification:
  - Many quantitative variables
  - Predict a categorical variable
Classifier

Attributes  Classifier  Label

(Demo)
Nearest Neighbor

How to classify a new individual:

- Find their **nearest neighbor**: the individual closest to them in the data set
- Assign the new individual the **same** label as that nearest neighbor
What is the type of `train_nn_su_classifier`?

- int or float
- array
- table
- function
- None of the above
<table>
<thead>
<tr>
<th>What is the type of train_nn_su_classifier(hgb_glc)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>int or float</td>
</tr>
</tbody>
</table>
What is the type of `train_nn_su_classifier(hgb_glc)
(make(array(15,130))))`?

- int or float
- array
- table
- function
- None of the above
Multiple Neighbors

- If data are noisy, asking just the closest neighbor might not be ideal for accuracy
- Instead, ask the $k$ closest neighbors, and take the majority label
Multiple Attributes

- We’ve used 2 attributes so far
- But nothing special about 2, just have to compute distances in higher dimensional spaces

(Demo)