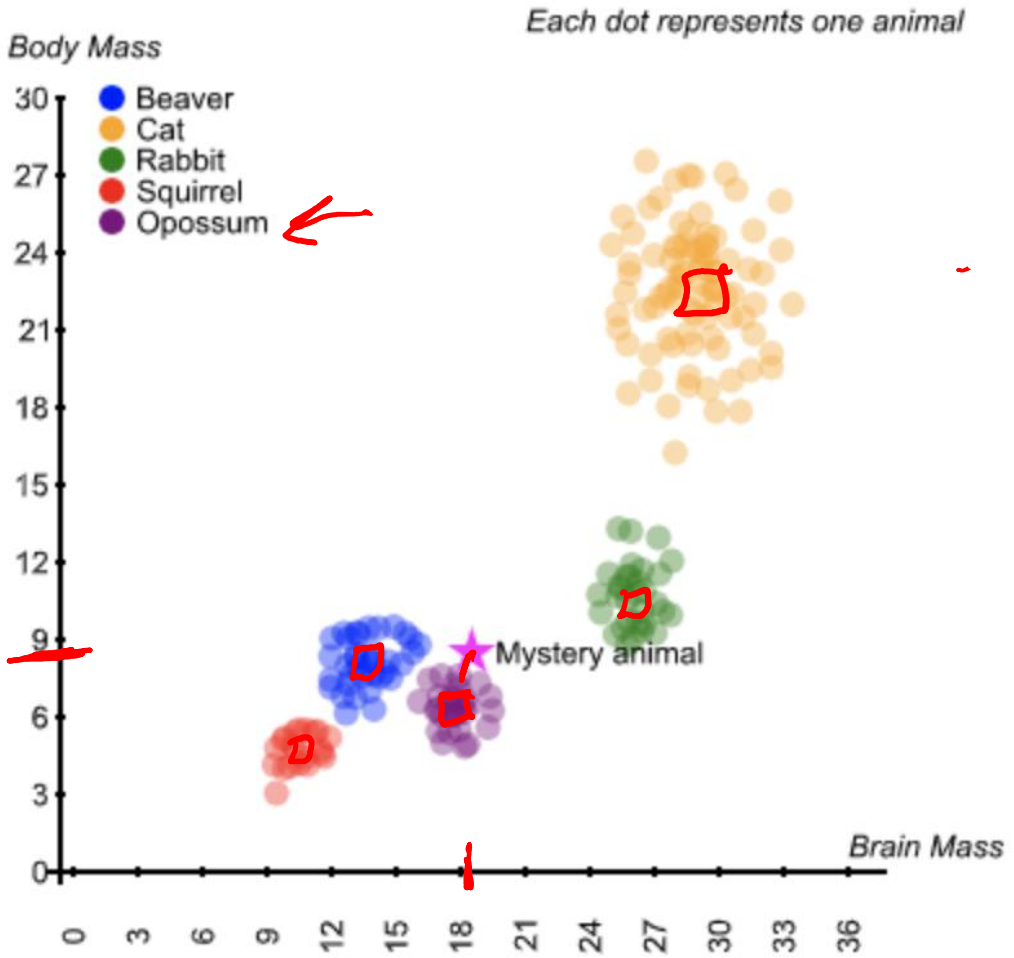
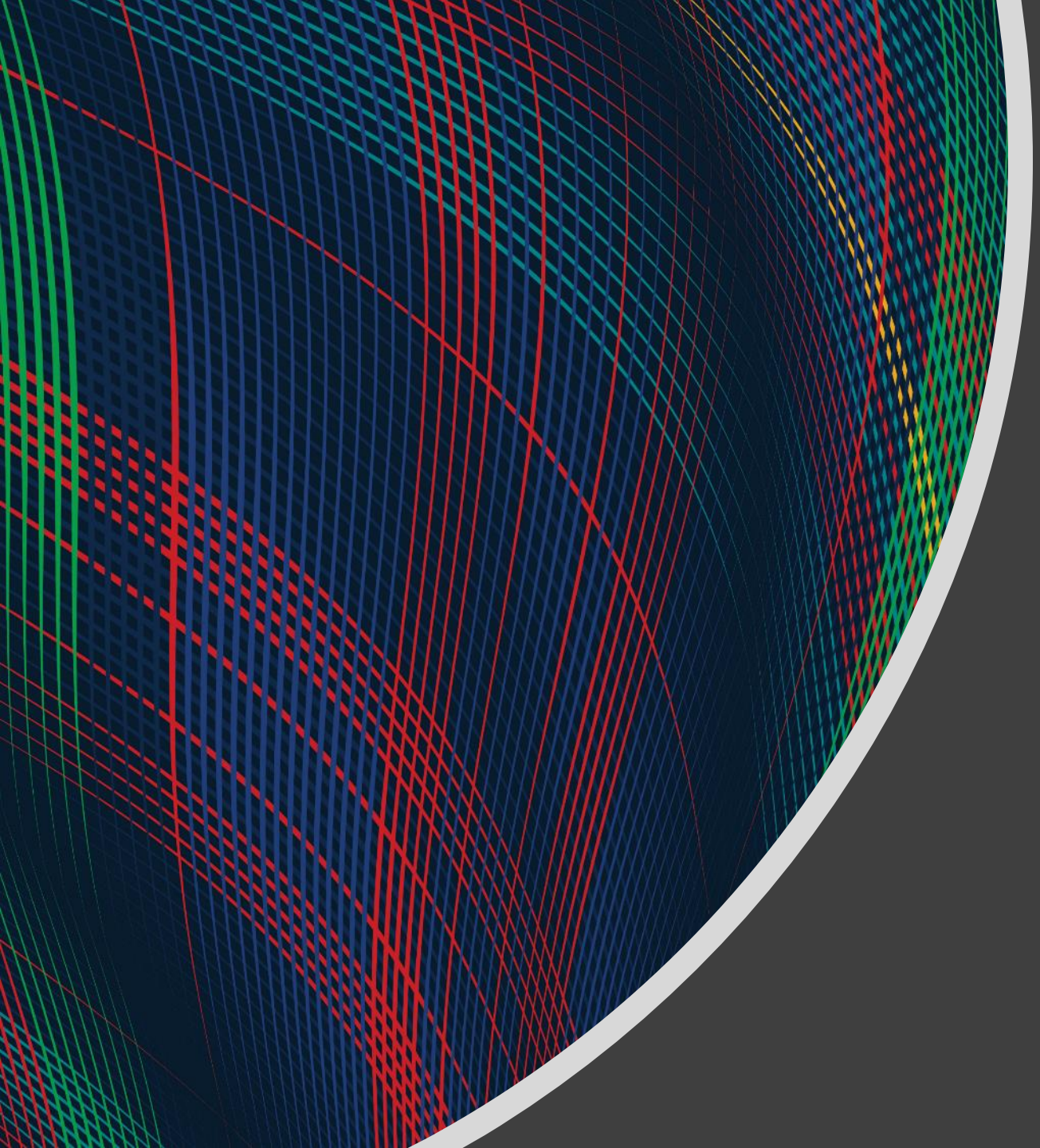
Visualizing Data to Make Predictions

What label should we assign the mystery point?





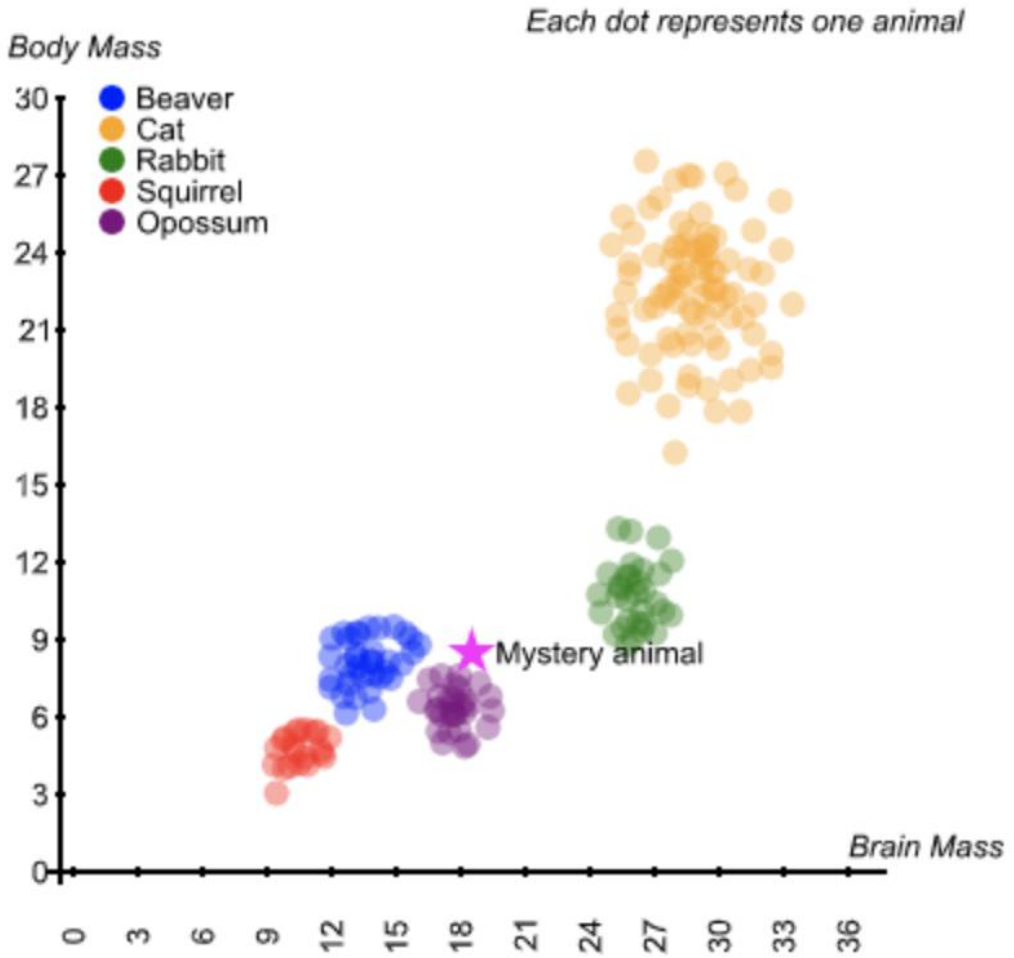
Demystifying AI

Nearest Neighbor

Instructor: Pat Virtue

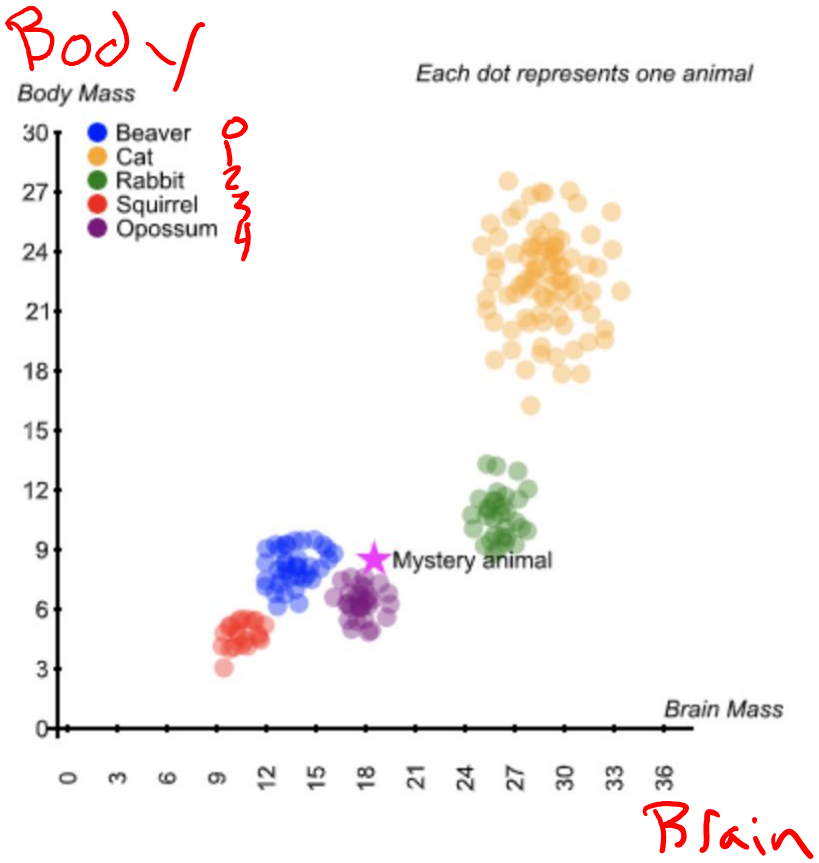
Visualizing Data to Make Predictions

What label should we assign the mystery point?



Input and Output?

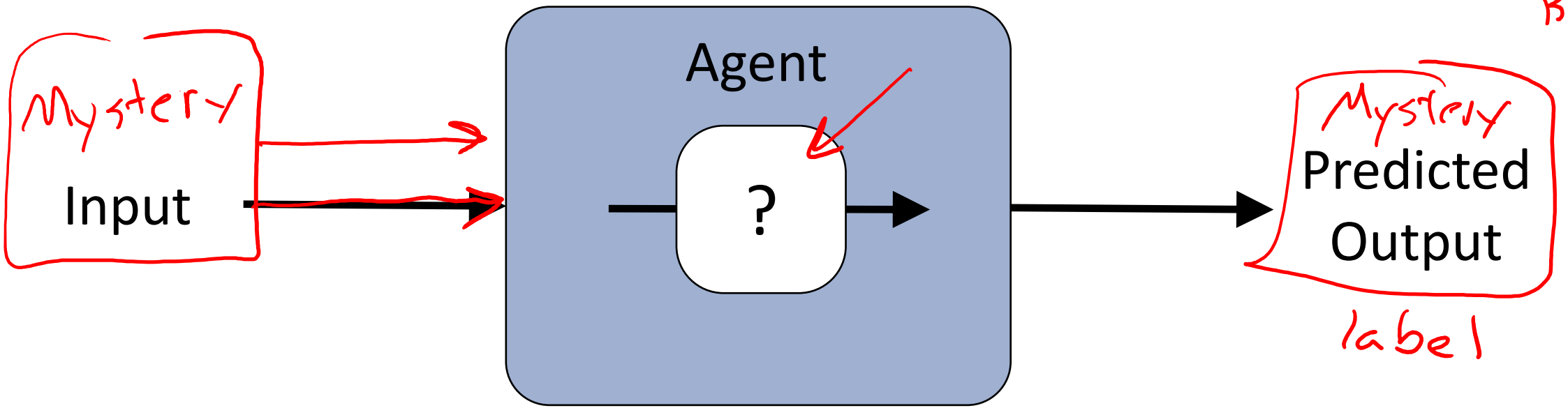
What is the input and the output for this task?



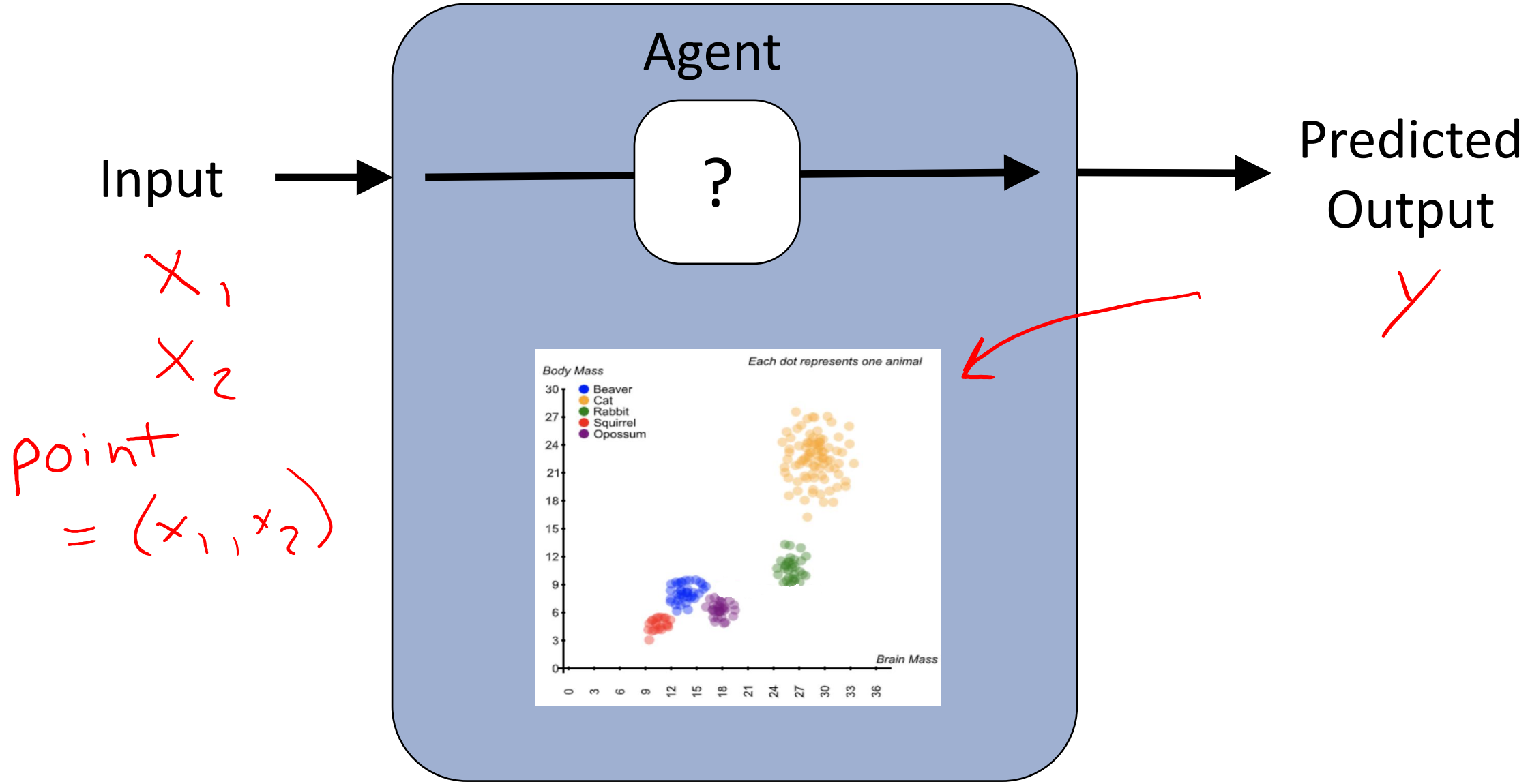
brain, body

predict(int , int)

return int

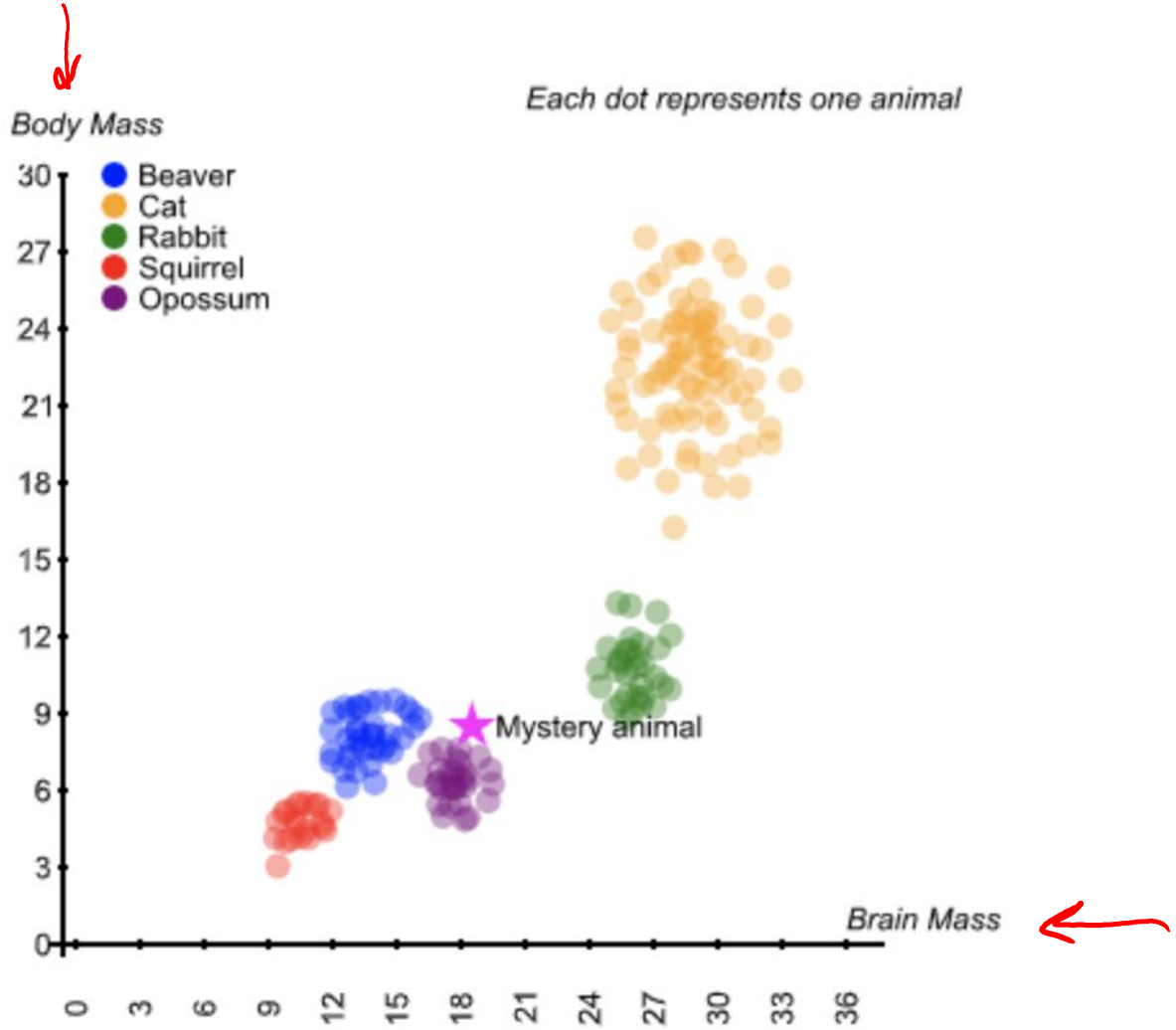


Agent Uses Data to Make Decisions

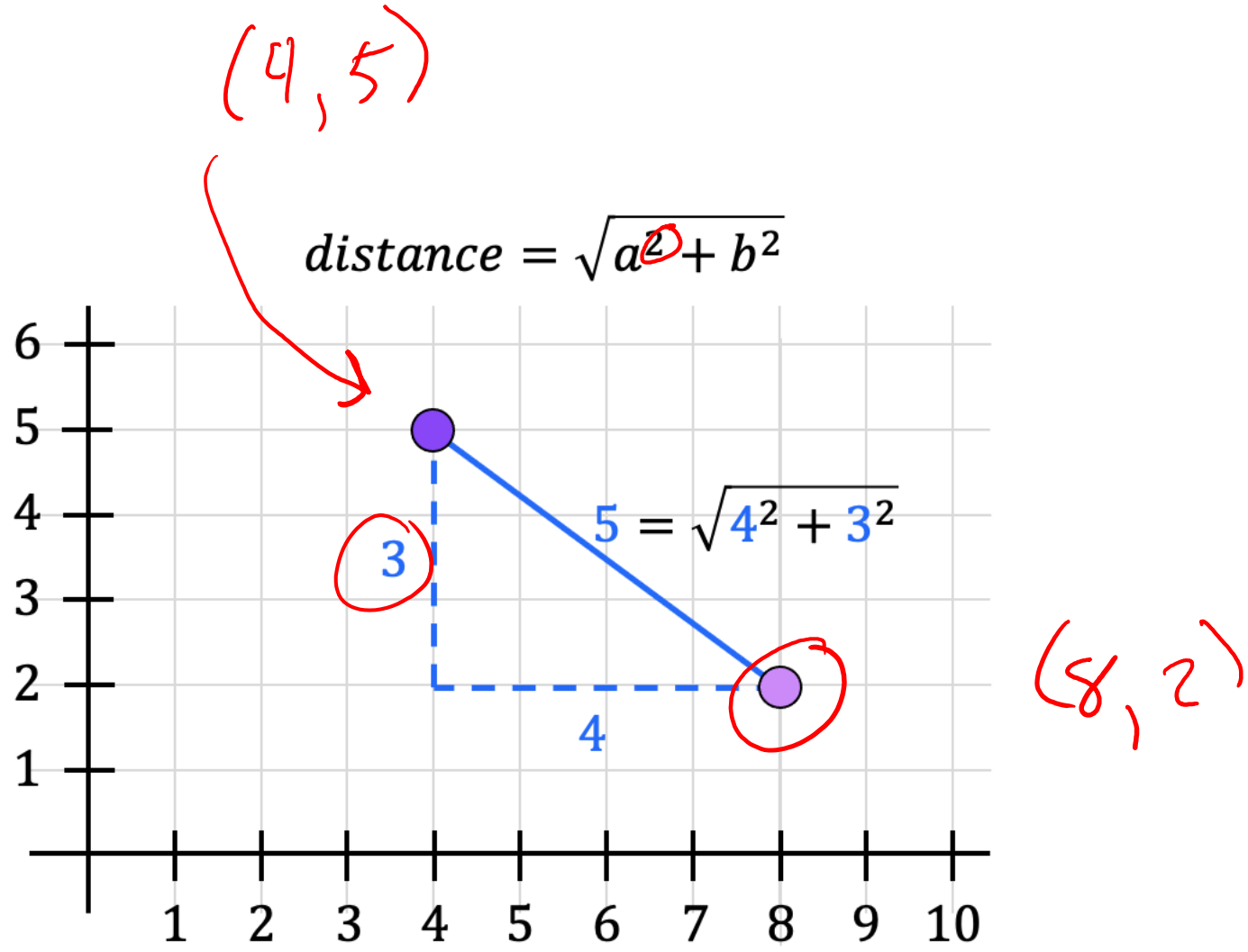


Using Distance as a Measure of Similarity

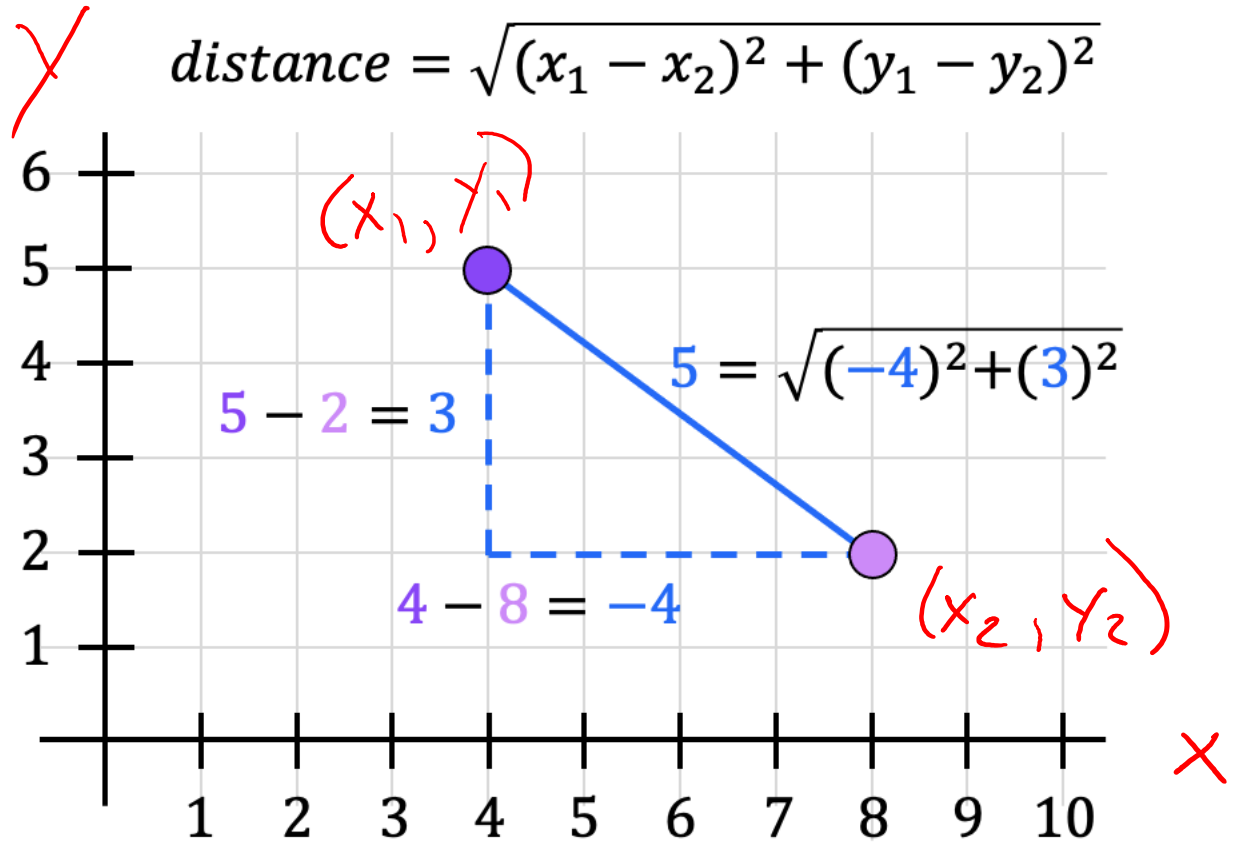
Assumption: closer points are more likely to be in the same category



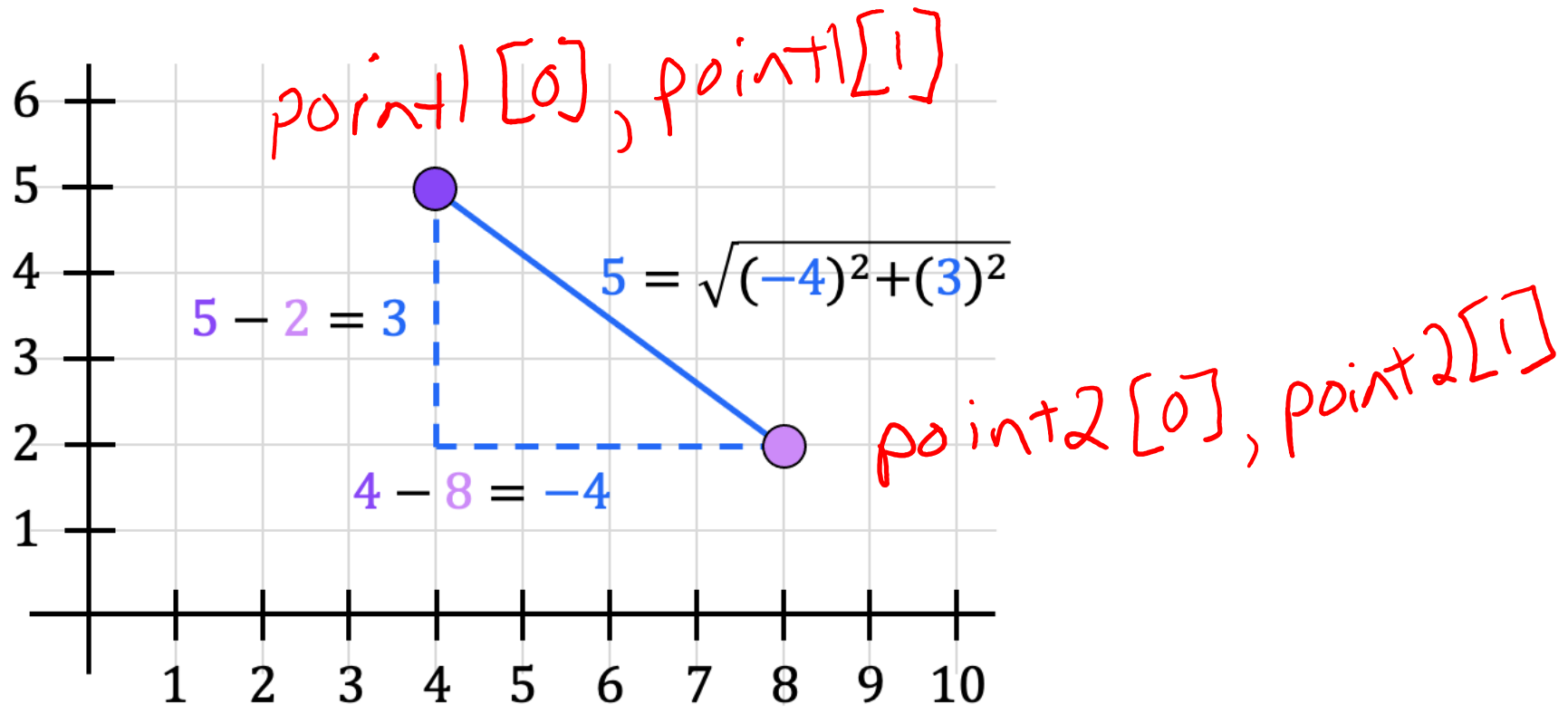
Distance



Distance

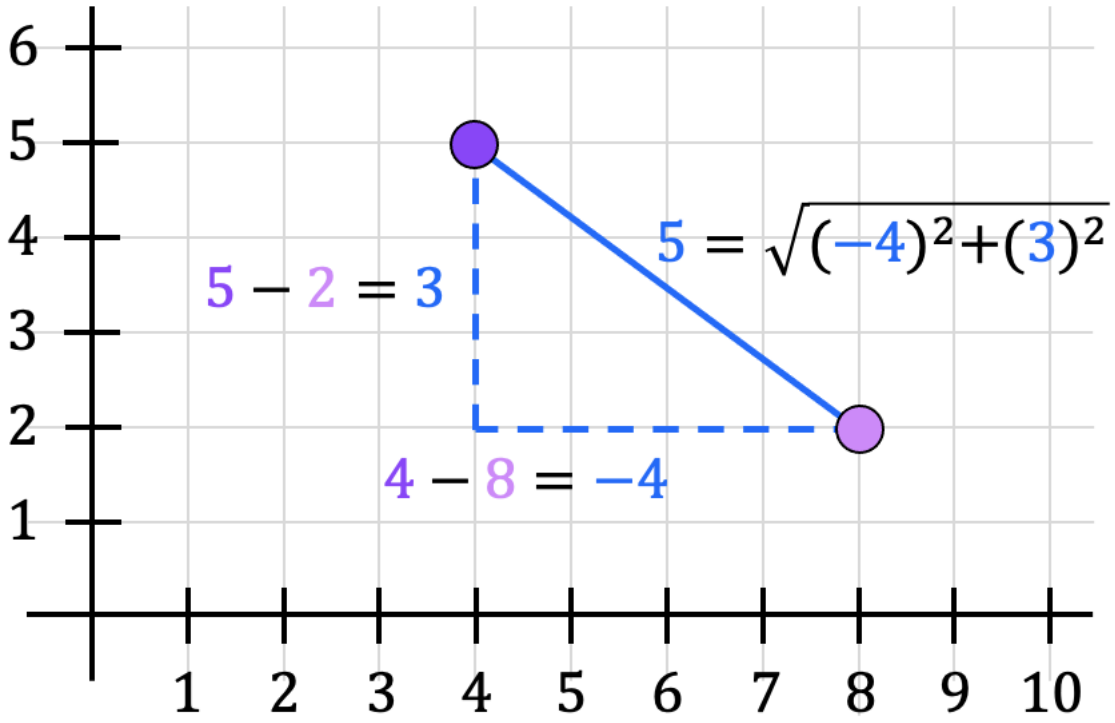

$distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
y
x
(x_1, y_1)
(x_2, y_2)
$5 - 2 = 3$
$4 - 8 = -4$
$5 = \sqrt{(-4)^2 + (3)^2}$
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6

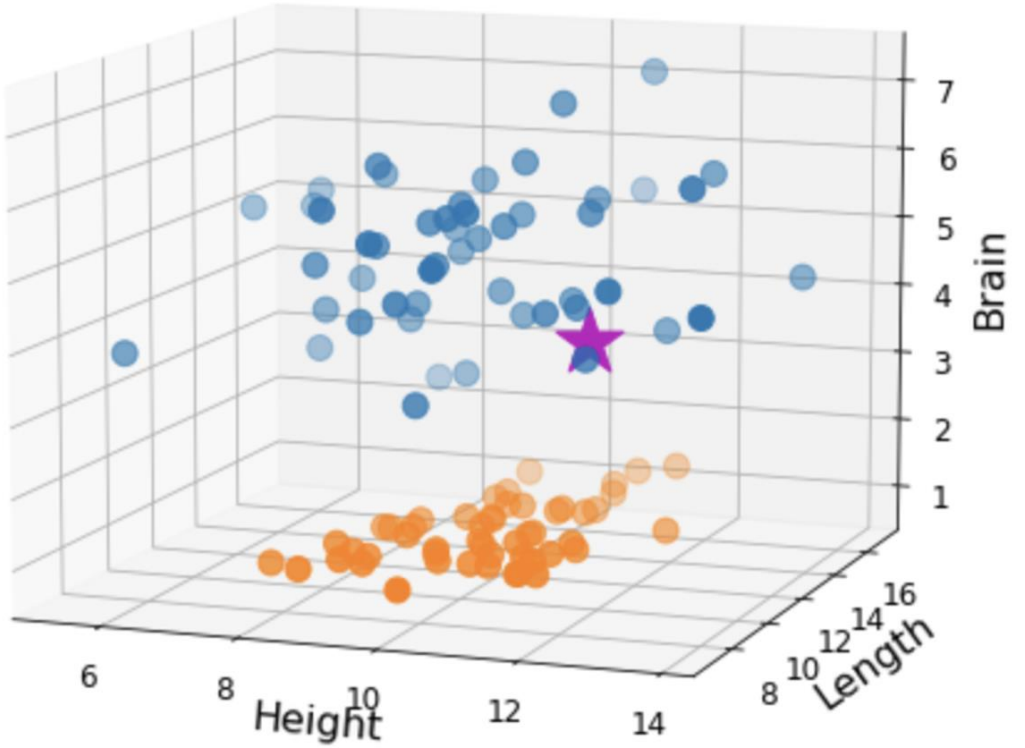
Distance


point1[0], point1[1]
$5 = \sqrt{(-4)^2+(3)^2}$
$5 - 2 = 3$
$4 - 8 = -4$
point2[0], point2[1]
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10

Distance

$$distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$



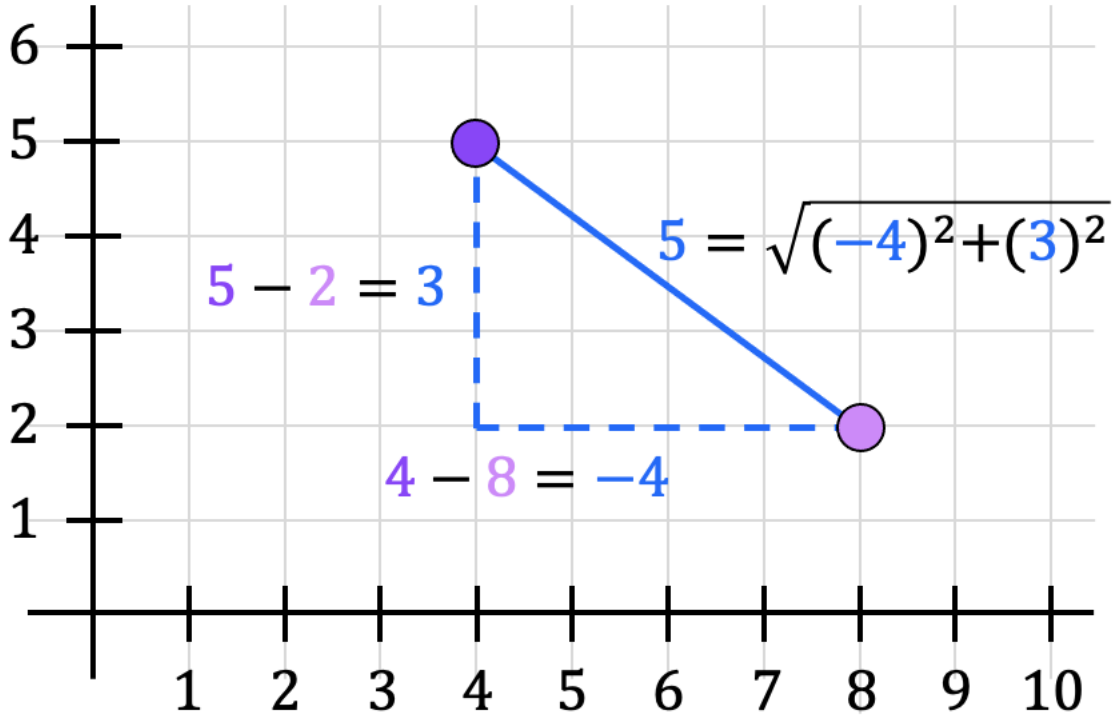


2D: $distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

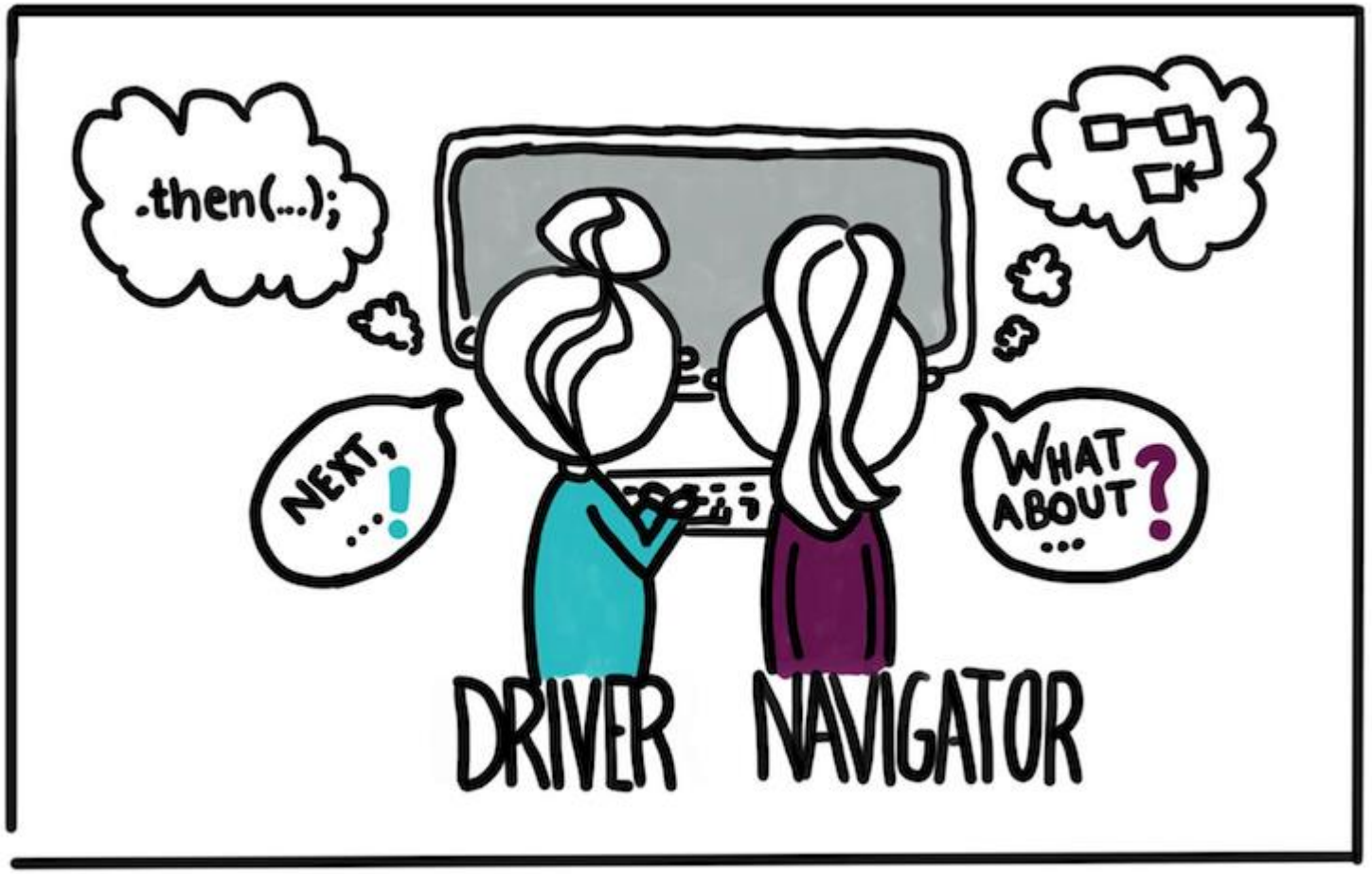
3D: $distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 + (w_1 - w_2)^2}$

Distance

$$distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$



Pair Programming



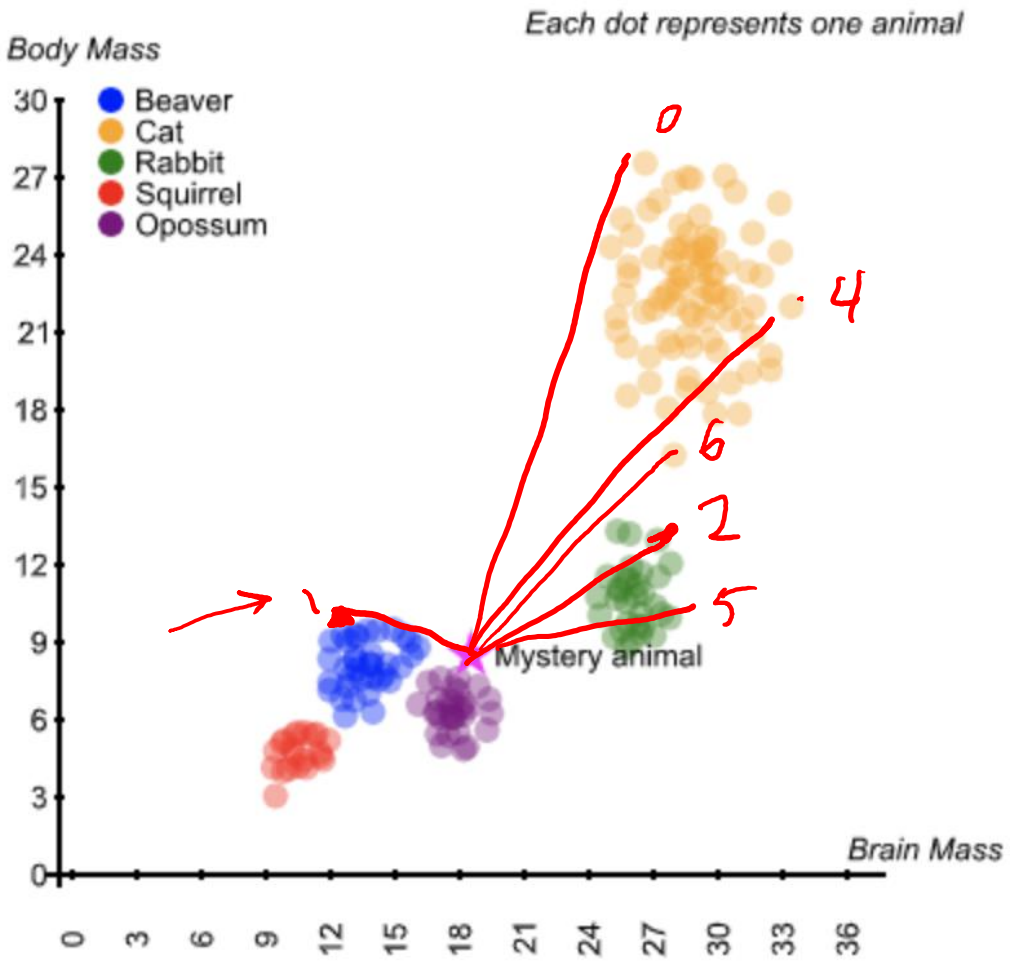
https://devopedia.org/pair-programming

Exercise 1: Distance

Shared Drive Folder/Lec5_NearestNeighbor

https://drive.google.com/drive/folders/11aAngUK5sifAnK4izULqPC-Y0QnDfbqG

Nearest Neighbor Algorithm



y_data

```
predict(x_data, x_new)
    best_dist = 1000000
    best_index = -1
    for each point
        dist to xnew
        if dist < bestdist
            update best

    return
```

Exercise 2: Nearest Neighbor

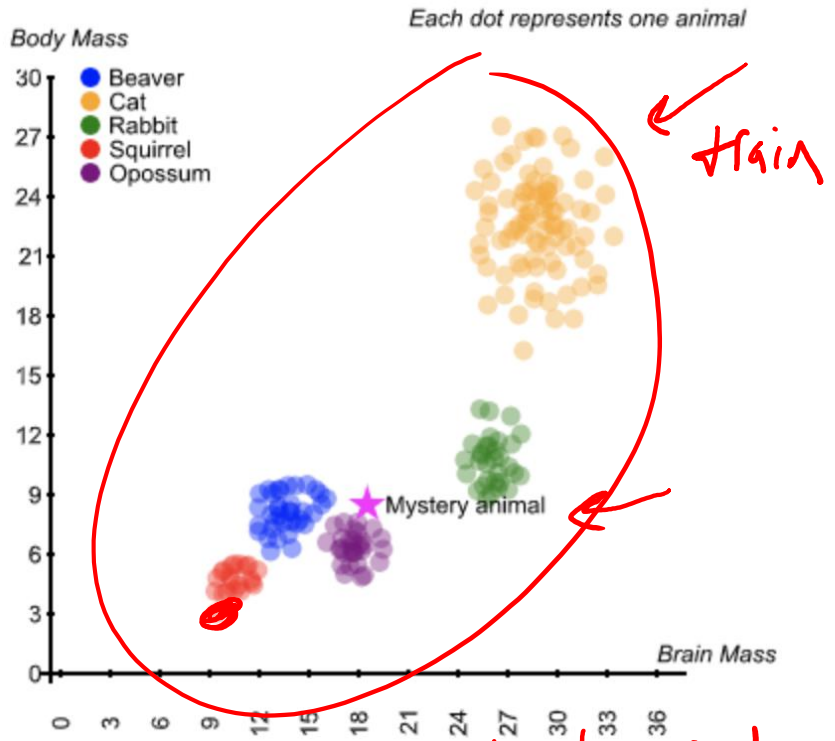
Shared Drive Folder/Lec5_NearestNeighbor

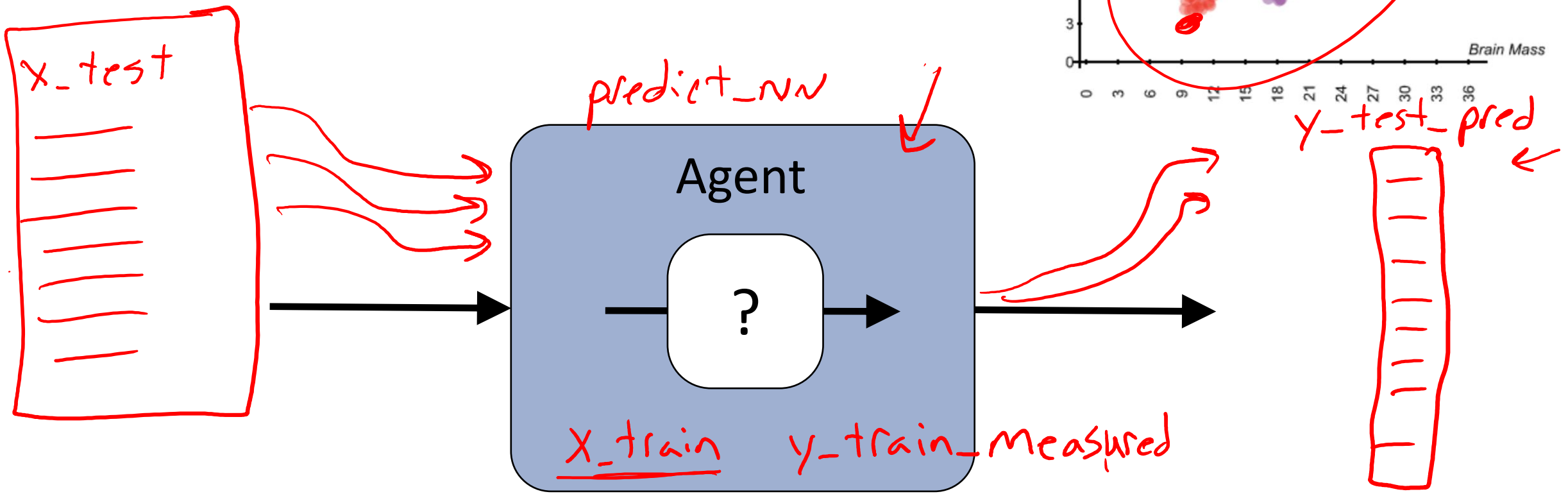
https://drive.google.com/drive/folders/11aAngUK5sifAnK4izULqPC-Y0QnDfbqG

Performance Measure

How well did we do?

test dataset x , y_measured





Performance Measure

- Classification error rate

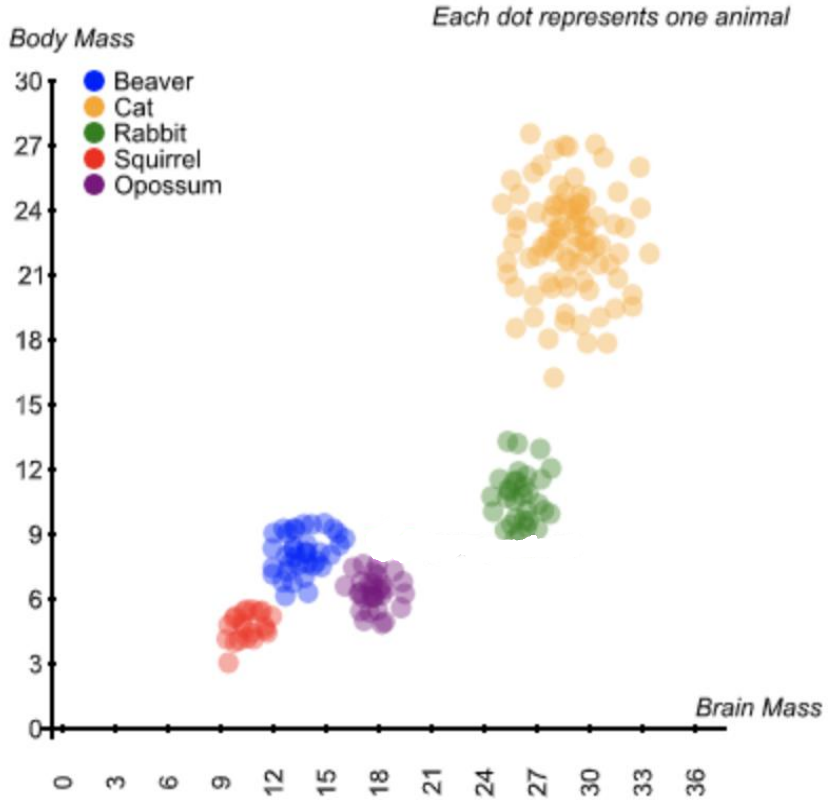
test
y_pred

0
3
2
2
0

test
y_meas

0
1
2
2
4

$$\frac{\text{num error}}{\text{num}}$$



Exercise 3: Error rate

Shared Drive Folder/Lec5_NearestNeighbor

https://drive.google.com/drive/folders/11aAngUK5sifAnK4izULqPC-Y0QnDfbqG