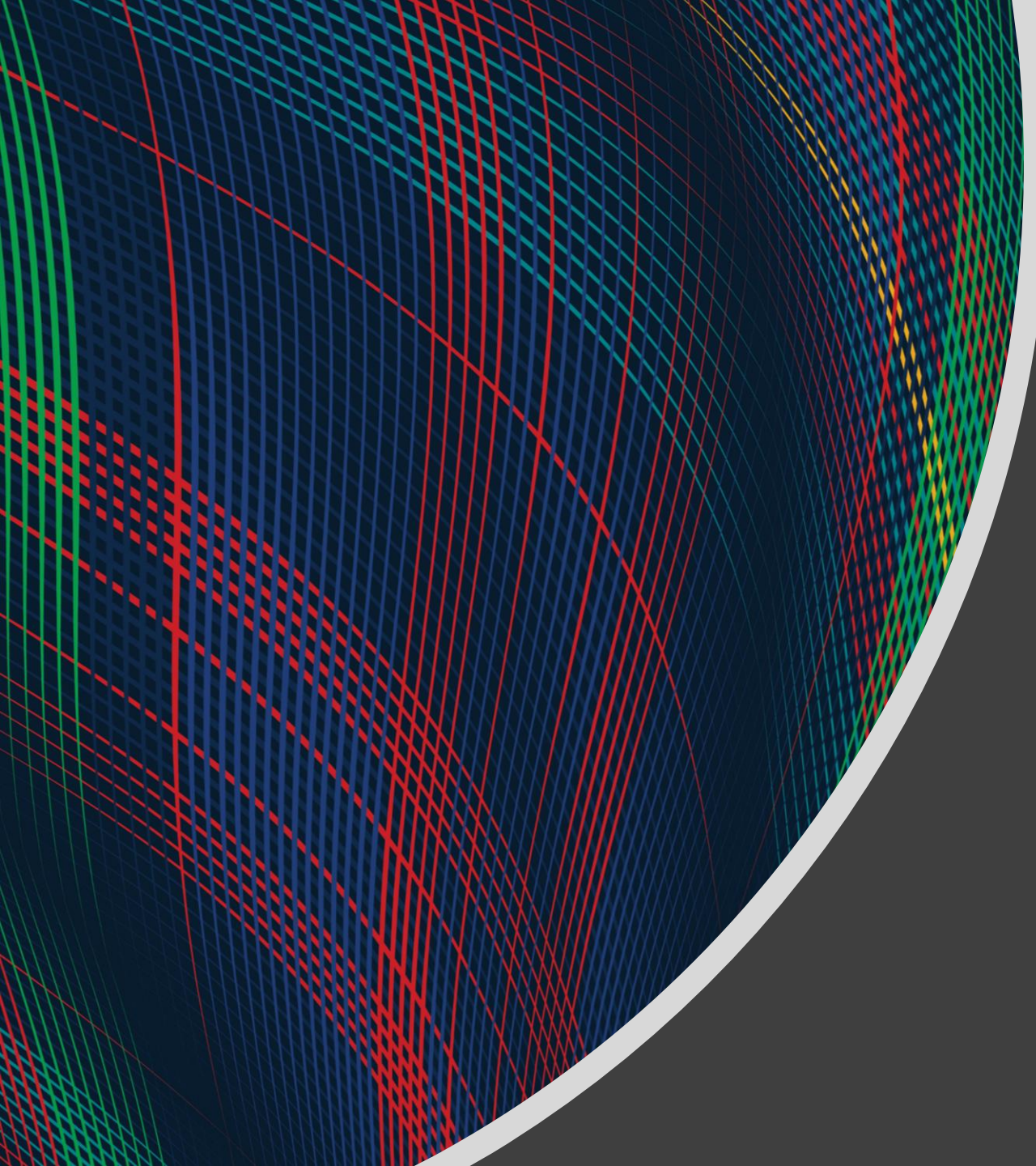


Announcements

Quiz

Homework

Exam



10-607
Computational
Foundations for
Machine Learning

Wrap-up Workshop

Instructor: Pat Virtue

Plan

Wrap up Bayes Nets and Variable Elimination

Nearest Neighbor Workshop

- Quick nearest neighbor intro
- Implement nearest neighbor in groups

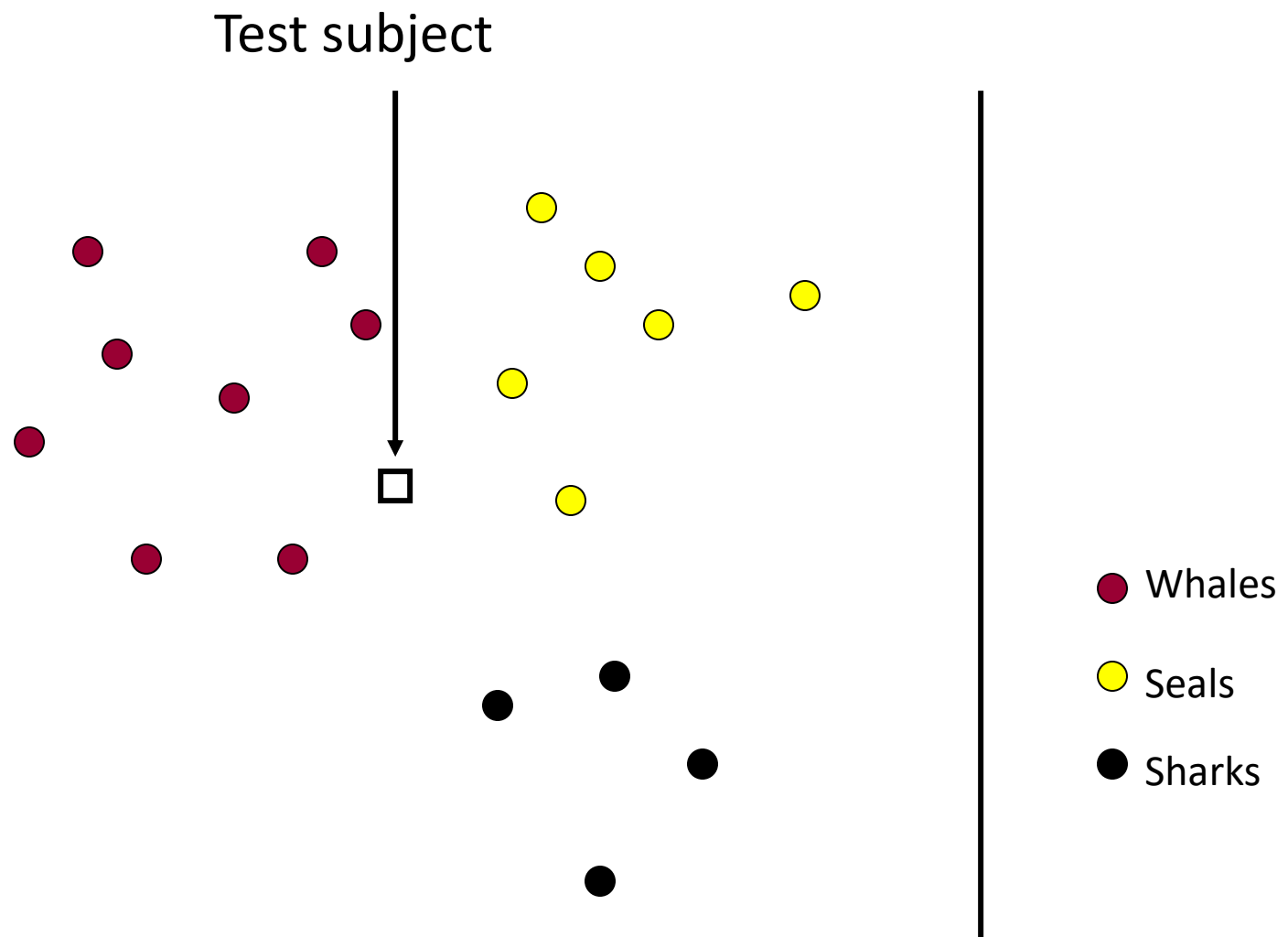
If time

- Pytorch tutorials

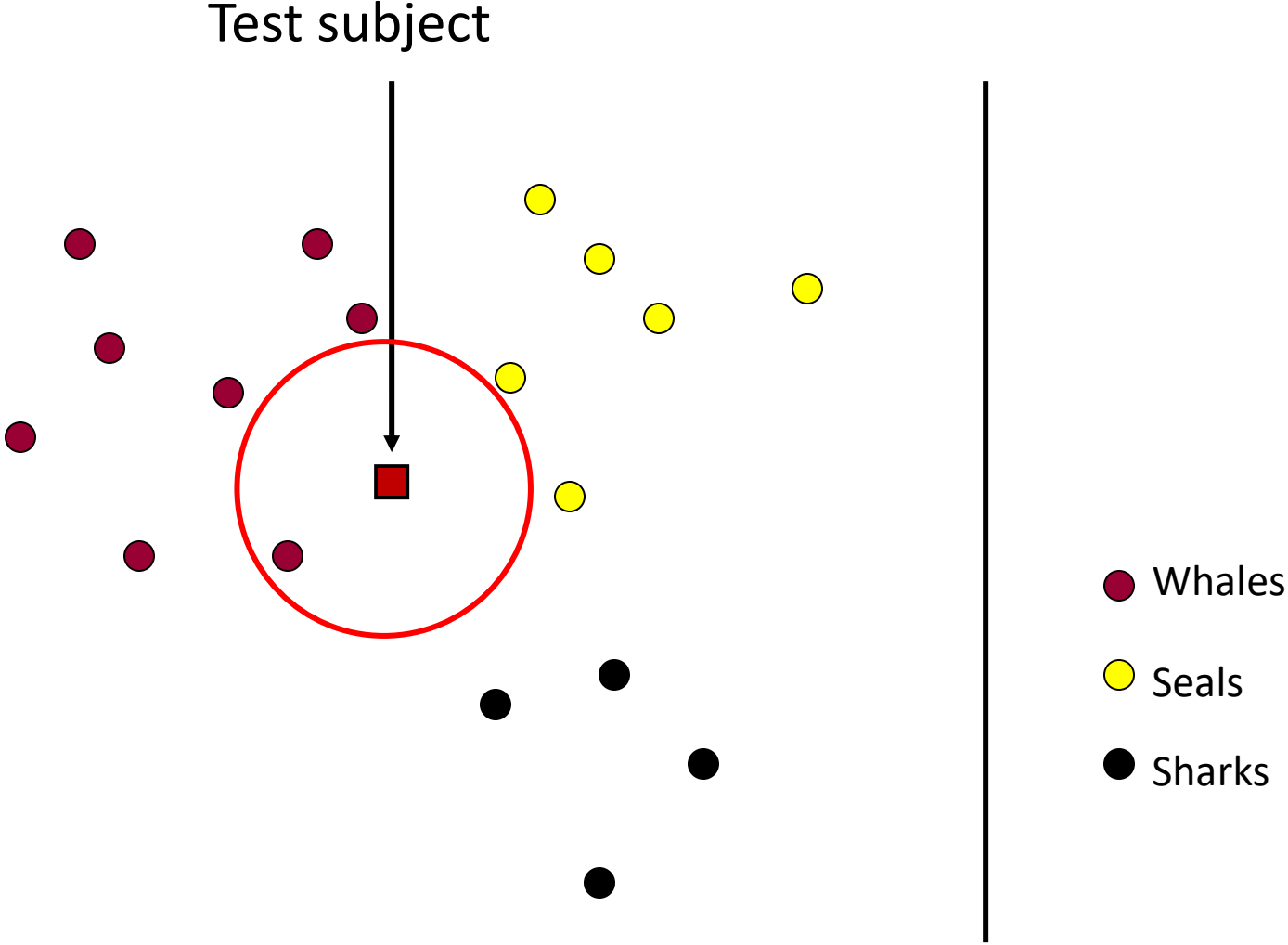
Bayes Nets

Previous lecture slides

Nearest Neighbor Classifier



Nearest Neighbor Classifier



Nearest Neighbor Classification

Given a training dataset $\mathcal{D} = \{y^{(n)}, \mathbf{x}^{(n)}\}_{n=1}^N$, $y \in \{1, \dots, C\}$, $\mathbf{x} \in \mathbb{R}^M$

and a test input \mathbf{x}_{test} , predict the class label, \hat{y}_{test} :

1) Find the closest point in the training data to \mathbf{x}_{test}

$$n = \underset{n}{\operatorname{argmin}} d(\mathbf{x}_{test}, \mathbf{x}^{(n)})$$

2) Return the class label of that closest point

$$\hat{y}_{test} = y^{(n)}$$

Need distance function! What should $d(\mathbf{x}, \mathbf{z})$ be?

Fisher Iris Dataset

Fisher (1936) used 150 measurements of flowers from 3 different species: Iris setosa (0), Iris virginica (1), Iris versicolor (2) collected by Anderson (1936)

| Species | Sepal Length | Sepal Width | Petal Length | Petal Width |
|---------|--------------|-------------|--------------|-------------|
| 0 | 4.3 | 3.0 | 1.1 | 0.1 |
| 0 | 4.9 | 3.6 | 1.4 | 0.1 |
| 0 | 5.3 | 3.7 | 1.5 | 0.2 |
| 1 | 4.9 | 2.4 | 3.3 | 1.0 |
| 1 | 5.7 | 2.8 | 4.1 | 1.3 |
| 1 | 6.3 | 3.3 | 4.7 | 1.6 |
| 1 | 6.7 | 3.0 | 5.0 | 1.7 |

Fisher Iris Dataset

Fisher (1936) used 150 measurements of flowers from 3 different species: Iris setosa (0), Iris virginica (1), Iris versicolor (2) collected by Anderson (1936)

| Species | Sepal Length | Sepal Width |
|---------|--------------|-------------|
| 0 | 4.3 | 3.0 |
| 0 | 4.9 | 3.6 |
| 0 | 5.3 | 3.7 |
| 1 | 4.9 | 2.4 |
| 1 | 5.7 | 2.8 |
| 1 | 6.3 | 3.3 |
| 1 | 6.7 | 3.0 |

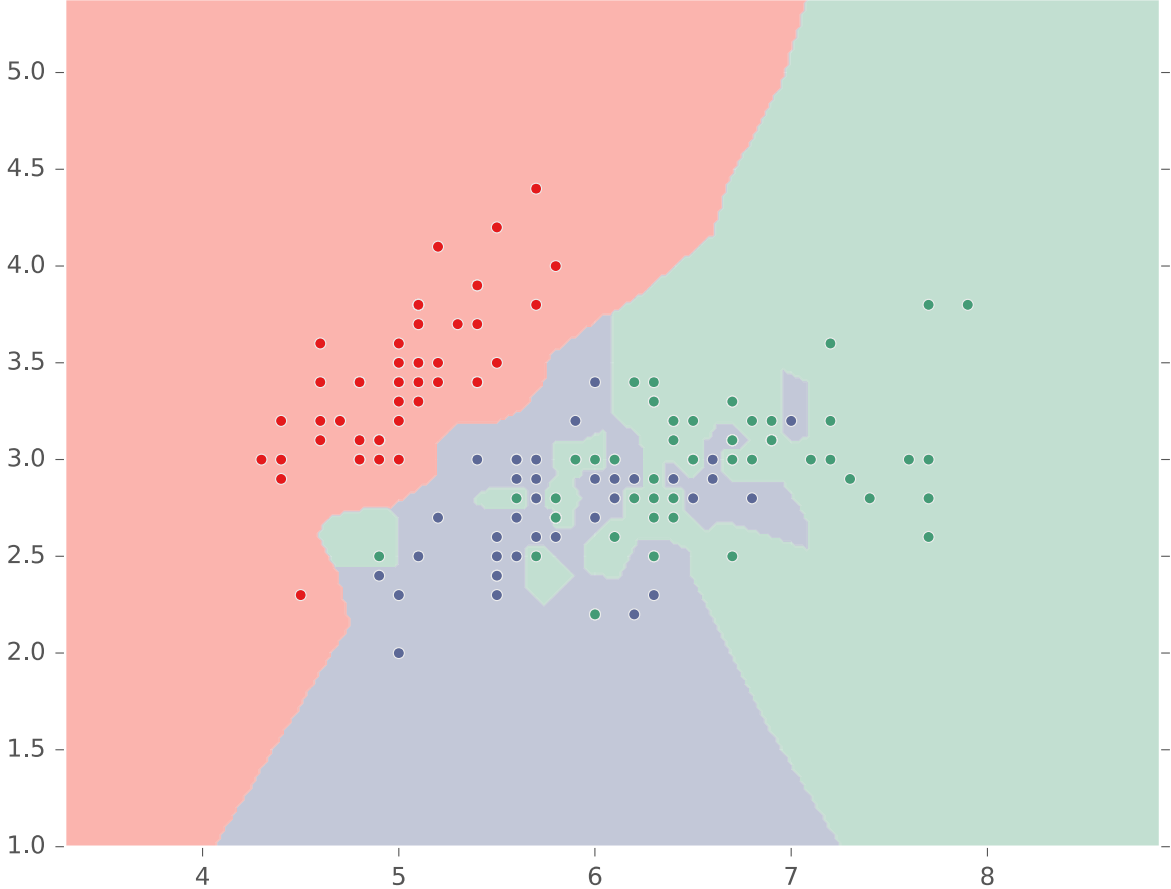
Deleted two of the four features, so that input space is 2D



Nearest Neighbor on Fisher Iris Data



Nearest Neighbor on Fisher Iris Data



Workshop: Nearest Neighbor

Task 1: Load into python and print contents

<https://www.cs.cmu.edu/~10607/data/animals1.csv>

Task 2: Plot animals1 data in 2D colored by label

Task 3: Implement nearest neighbor and predict label (\hat{y}) for
new_animal_body = 18.0
new_animal_brain = 8.5

Task 4: Run nearest neighbor with training data:

https://www.cs.cmu.edu/~10607/data/animalsND_train.csv

and measure the error rate on

https://www.cs.cmu.edu/~10607/data/animalsND_test.csv

Task 5: Calculate the confusion matrix for the animalsND_test dataset