

INTRO TO CLASSIFICATION

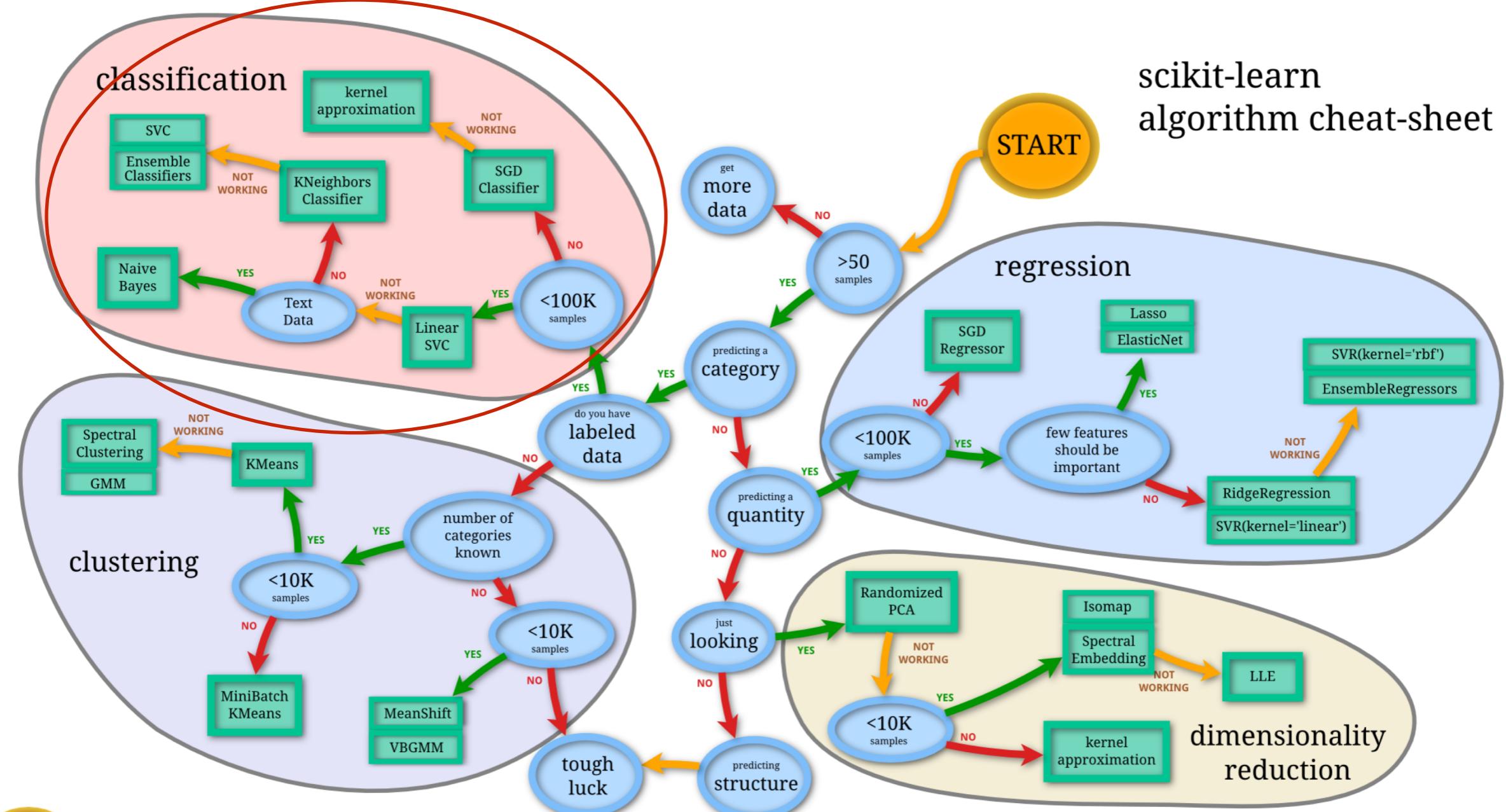
INTRO TO CLASSIFICATION

LEARNING OBJECTIVES

By the end of this lesson, we will be able to:

- Define class label and classification
- Compare and contrast regression and classification
- Code a basic classifier using the Iris dataset
- Build a K-Nearest Neighbors classifier using the sci-kit-learn library

scikit-learn algorithm cheat-sheet



Back

scikit
learn

INTRO TO CLASSIFICATION

- ▶ Any class guesses?

ACTIVITY: KNOWLEDGE CHECK



IN SMALL GROUPS, ANSWER THE FOLLOWING QUESTIONS

1. How have we been using linear regression?
2. How have we been evaluating our linear regression models?
3. Do you think we can use linear regression to identify gender?
What about hair color? Education level?
4. Would the same methods and metrics apply to the kind of
problem in (2.)?

DELIVERABLE

As a group, come to a consensus. Each group will answer one of these questions.

INTRODUCTION

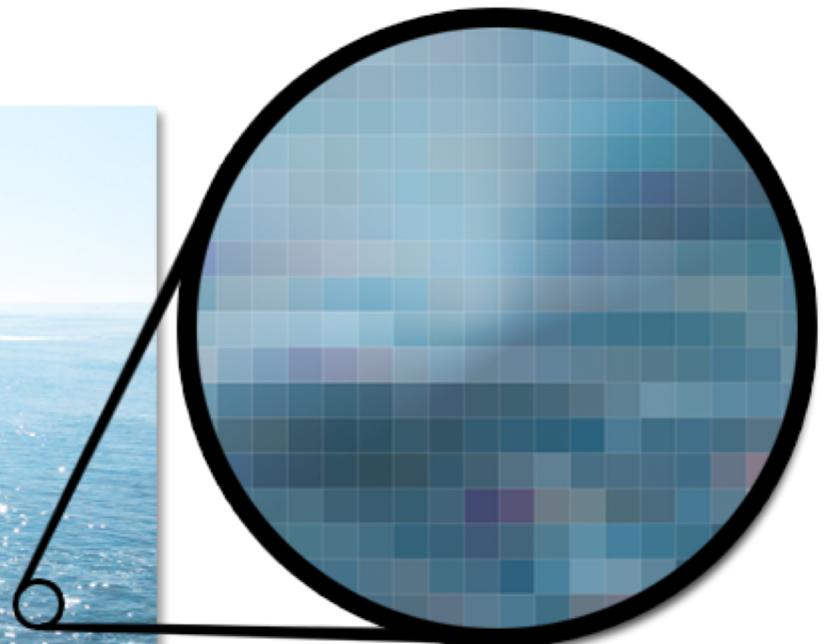
WHAT IS CLASSIFICATION?

WHAT IS CLASSIFICATION?

- ▶ **Classification** is a machine learning problem for solving a set value given the knowledge we have about that value.
- ▶ Many classification problems are trying to predict *binary* values.
- ▶ For example, we may be using patient data (medical history) to predict whether the patient is a smoker or not.

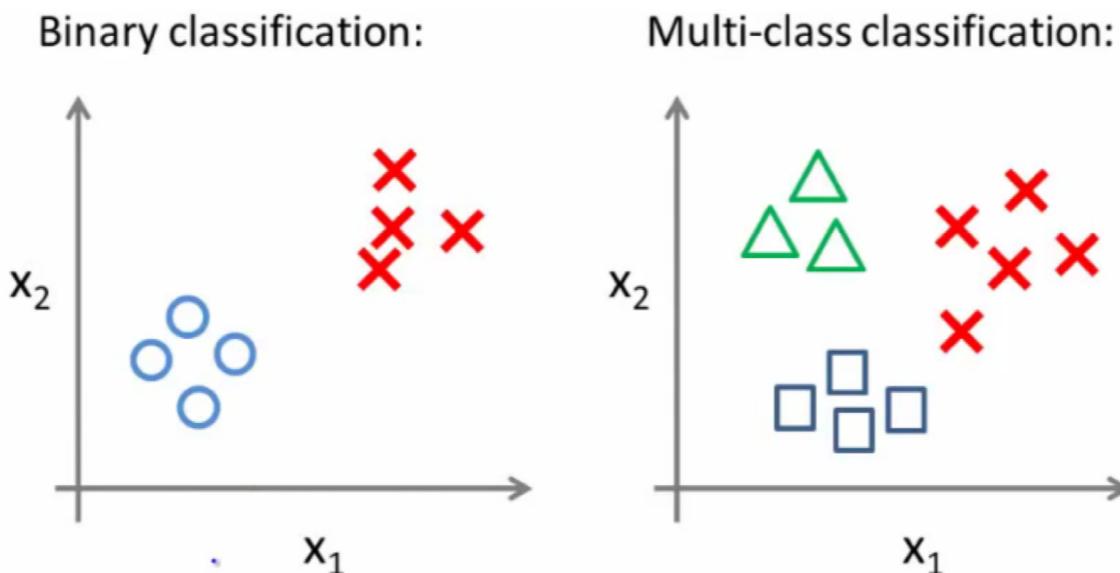
WHAT IS CLASSIFICATION?

- Some problems don't appear to be binary at first glance.
- What if you are predicting whether an image pixel will be red or blue?
- This is similar to the concept of dummy variables.



WHAT IS CLASSIFICATION?

- Binary classification is the simplest form of classification.
- However, classification problems can have multiple *class labels*.
- Instead of predicting whether the pixel is red or blue, you could predict whether the pixel is red, blue, or green.



WHAT IS A CLASS LABEL?

- ▶ A **class label** is a representation of what we are trying to predict: our *target*.
- ▶ Examples of class labels from before are:

Data Problem	Class Labels
Patient data problem	is smoker, is not smoker
pixel color	red, blue, green

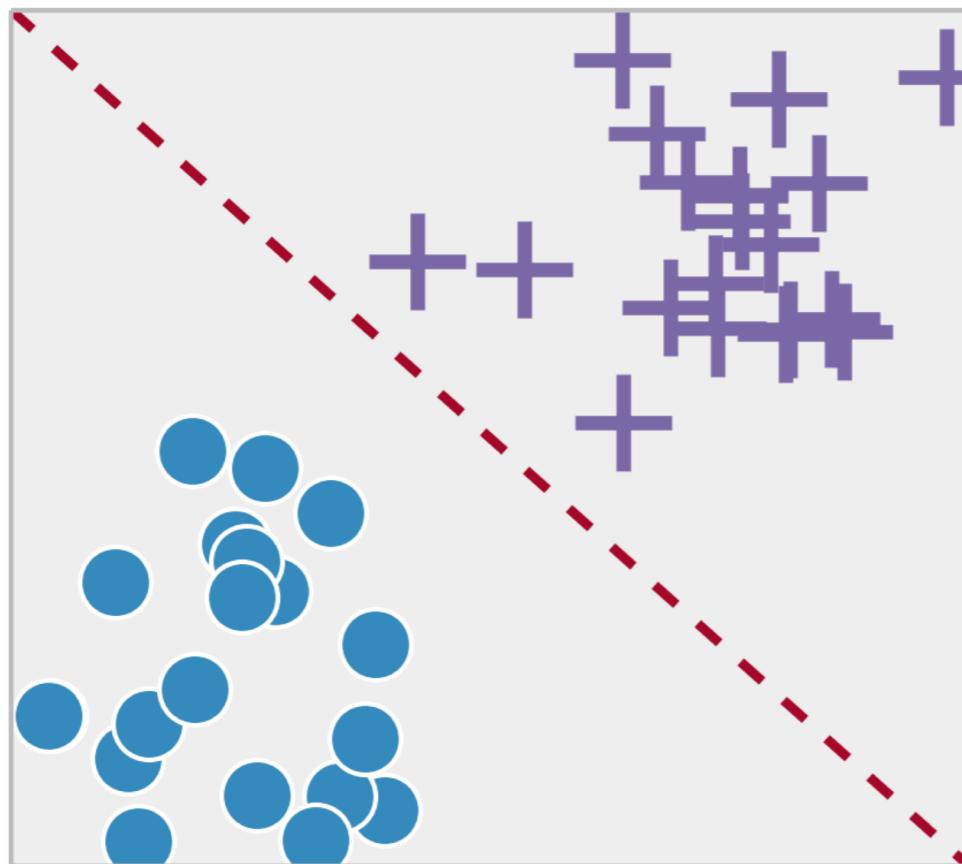
DETERMINING REGRESSION OR CLASSIFICATION

- ▶ One of the easiest ways to determine if a problem is regression or classification is to determine if our *target* variable can be ordered mathematically.
- ▶ For example, if predicting company revenue, \$100MM is greater than \$90MM. This is a *regression* problem because the target can be ordered.
- ▶ However, if predicting pixel color, red is not inherently greater than blue. Therefore, this is a *classification* problem.

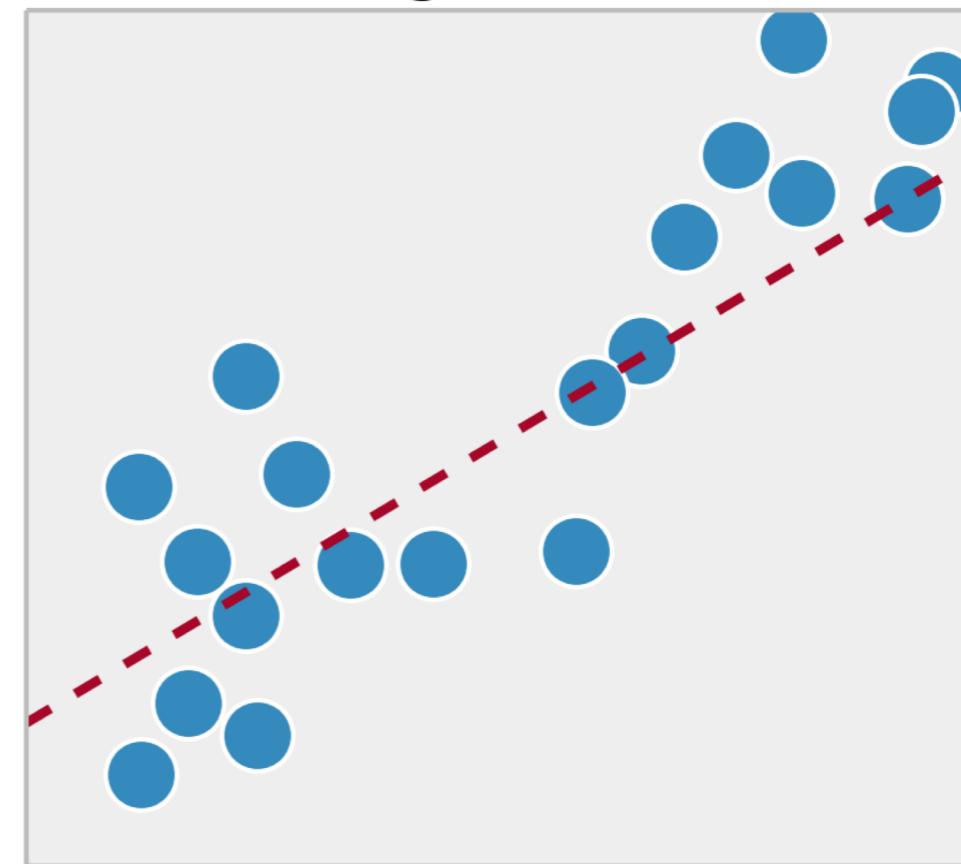
DETERMINING REGRESSION OR CLASSIFICATION

- Classification and regression differ in what you are trying to predict.

Classification



Regression



GUIDED PRACTICE

REGRESSION OR CLASSIFICATION?

GUIDED PRACTICE: GUESS THAT METHOD!



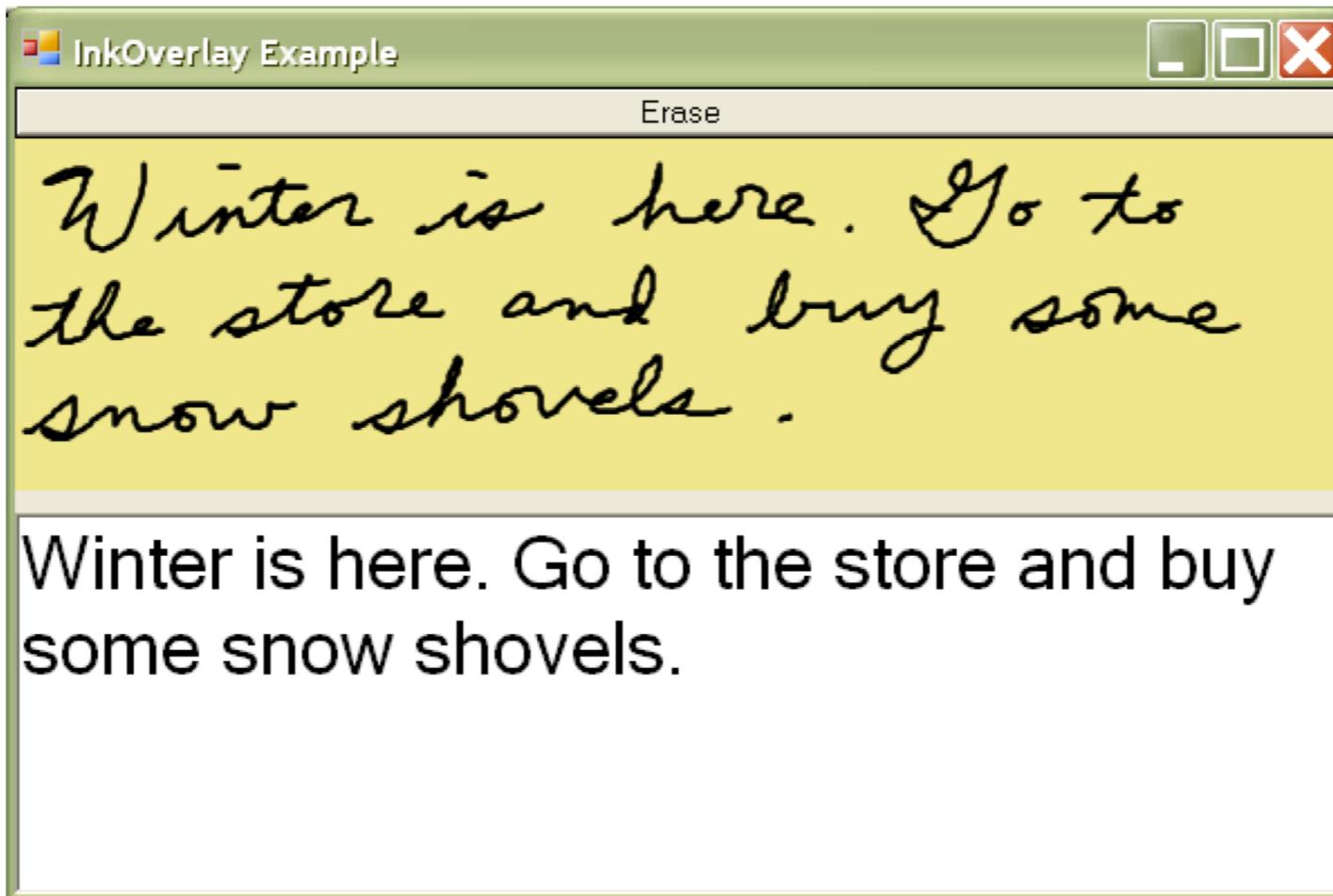
GUIDED PRACTICE: GUESS THAT METHOD!



GUIDED PRACTICE: GUESS THAT METHOD!



GUIDED PRACTICE: GUESS THAT METHOD!



ACTIVITY: REGRESSION OR CLASSIFICATION?

EXERCISE

DIRECTIONS

Review the following situations and decide if each one is a linear regression problem, a binary classification problem, a or multi-label classification problem, or neither:

1. Using the total number of explosions in a movie, predict if the movie is by JJ Abrams or Michael Bay.
2. Determine how many tickets will be sold to a concert given who is performing, where, and the date and time.
3. Using anonymized tweets, identify the sender
4. s, reduce the noisy signal from a living phone.
5. With customer data, determine if a user will return or not in the next 7 days to an e-commerce website.

DELIVERABLE

Answers to the above questions

**Let's jump on to the jupyter note book
now!**