

Lesson 5: Classification Models

45 minutes

Overview

In this lesson students will participate in an unplugged activity simulating one of the machine learning algorithms computers use to separate data into groups to help make decisions. Students will be tasked with helping a computer learn to classify food as fruits or vegetables, graph 20 different fruits on two axes comparing "sweetness" to "easy to eat", and then try to separate the data into groups - a fruit area, and a veggie area.

Question of the Day: How do computers learn to classify data?

Standards

Full Course Alignment

AI4K12 National Guidelines 2021

- ▶ **BI-3** - Computers can learn from data

CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis

Agenda

Warm Up (5 minutes)

Journal

Activity (35 minutes)

Creating Our Model

Testing Our Model

Wrap Up (5 minutes)

Journal

Objectives

Students will be able to:

- Explain how computers can separate data to make a decision

Preparation

- Review all materials for today's lesson.
- Cut out fruit & veggie cards - one set of cards for each group
- Ruler or straightedge for each student
- Check the "**Teacher's Lounge**" forum for verified teachers to find additional strategies or resources shared by fellow teachers

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the teachers

- **Classification Models** - Slides

▼ Make a Copy

For the students

- **Classification Models** - Activity Guide

▼ Make a Copy

- **Fruit & Veggie Cards** - Resource

▼ Make a Copy

Vocabulary

- **Categorical Data** - data that can be separated into groups
- **Classification** - predicting a category based on other features

Teaching Guide

Warm Up (5 minutes)

Journal

 **Prompt:** *How many colors are in the image below? Try finding “yellow”. Where does it start? Where does it end?*

Allow students to share their ideas with a neighbor before sharing with the class. Students will probably have different opinions about where yellow “starts” and “stops” - encourage several students to share their descriptions. It’s okay to not come to a consensus - the main purpose of the discussion is to motivate the following prompt.

 **Discuss:** *If you could teach a computer to separate the colors, how would you do it?*

Have a few students share their ideas with the class, building off of the language they used in the last prompt. There is no right answer here - it’s more important for students to generate ideas and practice describing a process to a computer.

 **Vocabulary:** Display the slide with today’s vocabulary on it:

- **Classification:** predicting a category based on other features.
- **Categorical Data:** data that can be separated into groups.

Remarks

Colors are an example of categorical data, which means they can be grouped into different categories like “yellow” or “green” or “blue”. When we ask a computer to look at a color and tell us what it is, we’re asking the computer to classify which color it belongs to. Classification is something we learn how to do with our senses, but computers must be taught to do. Today we’re going to investigate how computers can use classification to solve problems. We’re going to start with a simple example: determining if something is a fruit or a vegetable.

 **Question of the Day:** How do computers learn to classify data?

Activity (35 minutes)

Creating Our Model

 **Discuss:** Let’s say you go to a friend’s house and they make a meal that you’ve never eaten before, so you’re not sure if you’ll like it or dislike it. What are some qualities you look for in foods that you think help decide whether you’ll like it or dislike it?

Discussion Goal: This prompt helps seed ideas that we’ll return to later in the lesson. The goal is to brainstorm as many food qualities that students can think of. Some examples might include:

- how sweet the food is
- what color the food is
- whether it’s hot or cold
- whether it’s crispy or mushy
- how spicy the food is

It may be helpful to share some of your own answers as a way to jump-start the conversation with students.

Remarks

This is a great list! What we've basically come up with are different features of food - things we can use to help make decisions. Today, we're going to see if we can teach a computer to decide if a food is a fruit or vegetable. For now, we're going to focus on only two features: how sweet the food is, and how easy it is to eat. We've also got 20 foods that we can use to train the computer to help it tell the difference between fruits and veggies.

 **Distribute:** Pass out a copy of the Fruit or Veggie Worksheet to each student. Pass out a set of Fruit and Veggies Cards to each group of students. If possible, have the cards cut-out ahead of time to make it easier to share between students in the same group.

 Teaching Tip

Create a Class Graph: If you have the classroom space available, we recommend adapting this activity to be completed physically as an entire class on a graph that has been created on the floor of the classroom. The flow of the activity remains the same, except each student takes a card and physically stands in their recommended spot on the graph. The class then decides on how to use string to divide the students into "fruit" and "vegetable" groups. Any remaining students then take on the role of the "test" foods and see how they would be classified as either a fruit or vegetable.

 Assessment Opportunity

The responses in the activity guide can be used to formatively assess how well students have met the objectives of the lesson.

 **Display:** Show students the directions for this task and walk through them with students. Point out that students will need to use the food cards to create the points on their graphs.

Model: Graph at least one fruit or vegetable. Emphasize that students can use the recommended values, or they can use their own judgement to rate how sweet and easy to eat a food item is - this helps create slightly different graphs that can be compared in the next part of the activity.

 Teaching Tip

Graphing Skills: Consider this an opportunity to reinforce skills that students are learning in their math class, and be careful not to assume students are already fluent with graphing points. You may decide to model several points to help students graph their points, and you may encourage them to use the labels on the graphs to help place their points.

Circulate: Monitor students as they graph their points. For fruits and veggies that students have eaten before, encourage them to make their own decisions about how "sweet" and "easy to eat" the foods are. If a student hasn't heard of the food before, encourage them to use the recommended values instead.

 **Display:** Show students the example graph that you get when using only the recommended values. Point out that it appears that fruits and veggies seem to naturally group together. Ask students to raise their hands if their graphs also have fruits and vegetables clustered together.

 Teaching Tip

Imperfect Groups: since students have the option to use their own judgement on where to place their fruits, they may end up in a situation where some foods are intermixed with the opposite group. This is okay, and reflects what happens in the real-world where sometimes data is messy and can't be perfectly separated.

 **Remarks**

It looks like we picked two really great features to focus on - these are doing a really great job of separating our foods into fruits and veggies. This is really great for a computer - if it wants to determine whether a food is a fruit or a vegetable, it just needs to see where it is on the graph! Things get a little tricky in the middle of the graph though, so we need to make a decision: how can we divide our graph into a fruit side or a veggie side?

Model: Demonstrate on the slide how there are many different lines you could draw that separate the graph into a fruit-side and a veggie side. The most important factors to consider are that you have as many fruits as you can on one side, and as many vegetables as you can on the other. As a class, decide on a line you want to use to divide the example graph on the slides and draw it in.

 **Do This:** Have students draw their own line to split the graph into a fruits side and a veggies side.

Remarks

Congratulations! We just created a model that we can use to help a computer make a decision! When a computer sees a new food, it will plot a new point on the graph and use that to decide whether it's a fruit or a vegetable. Let's see how our models do with some test foods!

Content Corner

This activity simulates the **Support Vector Machine (SVM)** machine learning algorithm for making predictions based on data. This algorithm focuses on separating data while maximizing the margin of separation.

If you would like to learn more about SVM and other machine learning algorithms, ml-playground.com has an interactive widget and links to additional resources. This website is intended for adults looking to learn more about machine learning, especially considering the amount of math involved, so we do not recommend sharing this with students.

Testing Our Model

Display: The next several slides have different foods to test against the model. These examples have been designed to motivate certain conversations. Have students keep track of how many examples their models get correct and incorrect.

 **Display - Raspberry & Pumpkin:** These two examples are designed to match the expectations of the model. Most students' models should correctly identify the Raspberry as a fruit and the pumpkin as a vegetable.

 **Display - Tomato and Cucumber:** These two examples are more ambiguous and, even within your classroom, students may not know whether these are fruits or vegetables. Both of these items are technically fruits because they contain seeds, but are usually used as vegetables when making foods. Ask the class to share how their models classified these items, noting that students may have different answers depending on how they drew their lines.

Teaching Tip

Right on the Line: Depending on how your students draw their lines, they may have a situation where a test fruit is exactly on the line they drew. If this happens, remind students that the model has to make some kind of decision, so they need to come up with their own rule for what happens when a point lands right on the line - is it always a fruit? Or always a vegetable?

 **Discuss:** *Without looking at anyone else's paper, why do you think some people got different answers than others?*

Discussion Goal: Students should realize that how they decided to draw their line impacts the decision that their model makes. You can model this in the front of the room with the example graph on the slides - if the line becomes a little more tilted or goes completely vertical, it can change the way that some foods are classified. Emphasize that this is why testing is important - it helps make sure the decisions we're making are correct.

 **Display - Lemon and Sweet Potato:** These two examples are intentionally designed to be incorrect. Students will likely misclassify these two items - their model will think a lemon is a vegetable and that a sweet potato is a fruit.

Remarks

Most models will probably get this answer incorrect, which means our foods were misclassified. This happens in real-life too - sometimes models make mistakes and get incorrect answers. For something like classifying fruits and vegetables, it may not be a huge deal. But if we were making decisions about people instead, this could be a really big deal. This is why testing is so important - we need to make sure our models are doing a good job before letting them make decisions for us.

 **Display - Ice Cream and Flaming Hot Cheetos:** These two examples are deliberately silly and fun, and you should lean-in to how the model will likely classify Ice Cream as a fruit and Flaming Hot Cheetos as a vegetable.

Remarks

I think we can all agree these examples are pretty silly. They're an example of using a model for something other than what it was intended for. Since we trained our model only on fruits and vegetables, it doesn't work very well on other types of foods. If we wanted to improve our model, we'd probably need a lot more data and a lot more features - more than just how sweet and easy-to-eat food is. This gets a lot harder for us as humans to understand, and we couldn't use our graph paper anymore. Luckily, computers can help us with this! Tomorrow we'll see how a computer looks at lots of data and lots of features to help make decisions

Wrap Up (5 minutes)

Journal

 **Prompt:** *If you were to break down the process we went through today to create a model and use it into steps for a computer to follow, how would you describe those steps?*

Discussion Goal: This prompt bridges today's lesson to tomorrow's, where a computer will be doing most of this work behind-the-scenes. Read through student responses to see how well they understood the process of today. Answers could be as simple as:

- Graph the data
- Draw a line to separate the data into zones
- When new data comes, see which zone it lands in to make a decision



This work is available under a **Creative Commons License (CC BY-NC-SA 4.0)**.

If you are interested in licensing Code.org materials for commercial purposes **contact us**.