

Lesson 11: Numerical Models

45 minutes

Overview

In this lesson, students participate in an unplugged activity simulating a zombie outbreak. Students must predict which parts of town have the least amount of zombies using data from a neighboring town. Students will use degrees of similarity and averages to make predictions about the number of zombies at a particular location. Then, students are rescued and get to compare their predictions to the actual numbers as a way to discuss how accuracy is different for numerical data compared to categorical data.

Question of the Day: How do computers learn to make predictions with numerical 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)

Wrap Up (5 minutes)

Journal

Objectives

Students will be able to:

- Explain how accuracy calculation for numerical data is different from categorical data
- Explain how computers can make decisions by comparing similarities in data

Preparation

- Review all materials for today's lesson.
- Print or prepare to share online the activity guide - one per student or group.
- 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

- **Numerical Models** - Slides

▼ Make a Copy

For the students

- **Numerical Accuracy** - Activity Guide

▼ Make a Copy

- **Zombie Prediction** - Activity Guide

▼ Make a Copy

Teaching Guide

Warm Up (5 minutes)

Journal

 **Prompt:** *You all make plans to hang out at a park over the weekend.*

- *Zoey lives 12 minutes away from the park*
- *Isaac lives 24 minutes away from the park*
- *Hawa lives 5 minutes away from the park*
- *Nico lives 30 minutes away from the park*
- *Kim lives 13 minutes away from the park*

Aaron lives closest to Zoey, Isaac, and Nico. How long do you predict it will take Aaron to get to the park? Explain your reasoning.

Have students journal individually first, then share with a neighbor.

 **Share Out:** Have students share out their responses to both prompts.

Discussion Goal: Even though students can't be 100% sure about the answer, students should use the clues about the friends Aaron is "closest to" to help guide their answers. Students may have a wider variety of answers, but they should all rely on using the values from Zoey, Isaac, and Nico to help predict how long it will take Aaron to get to the park.

Remarks

In both of these situations, we can use information about Aaron's friends to help predict something about Aaron. It may not be 100% correct, but our prediction will probably be close enough. This is another way that computers can find patterns in data with machine learning - looking at similar data to make predictions. Today, we'll explore how machines can make decisions using this same technique.

 **Question of the Day:** How do computers learn to make predictions with numerical data?

Activity (35 minutes)

 **Distribute:** Pass out the **Zombie Prediction** activity guide to each student.

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 overview from the activity guide, which explains how there has been a zombie uprising and we are working to figure out safe places to hide.

 **Do This:** Have students look over the data showing how many zombies are at eight different locations. They should record their observations at the bottom of the activity guide.

Teaching Tip

Answer Key: An answer key to this activity is provided to verified teachers in the links section of the lesson plan

 **Share Out:** Have students share out their answers to the questions at the bottom of the activity guide.

Discussion Goal: Students may notice that zombies tend to group at loud, outdoor areas. Students may also notice zombies tend to group at locations where there are humans or animals, like schools or zoos. Students may try to justify this pattern by making references to zombies in pop culture, but it's okay if this doesn't come up.

 **Display:** Have students flip to the next page of the activity guide, which is also represented on a slide. This activity guides students through the process of turning the zombie data into a model that can make predictions about how many zombies are in a particular location.

Model: Help students track how similar each row of the table is similar to Location A. This involves counting each time a cell from Location A matches a cell in the table. For example, if the only thing both locations have in common is that they are indoors, then we would write a 1 for the similarities with Location A. Or, if both locations are indoors, loud, and have sidewalks: then we would write a 3 for the similarities with Location A.

Model: Help students predict how many zombies will be at Location A by finding the three most similar locations, then averaging the number of zombies at each location. It can be helpful to model this process for Location A, since students will repeat this process for Locations B and C on the back of the activity guide.

Content Corner

This activity simulates the **K-Nearest Neighbors (KNN)** machine learning algorithm for making predictions based on data. The **K** represents how many neighbors you look for - in this activity, we're using **K=3** because students find the three most similar data points to calculate the average.

If you would like to learn more about KNN 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.

 **Do This:** On the next page of the activity guide, students are given two new locations to predict the number of zombies. They will repeat this same process for these locations: determine the three most similar locations, then find the average to predict the number of zombies.

Circulate: Monitor students as they complete this process. An answer key is provided to help check answers.

 **Share Out:** Ask students to share out only their predictions for locations B and C, and which location has the least number of zombies.

Remarks

This is an example of how we can take new data and compare it to our existing data, then use those similarities to make predictions. This happens a lot in machine learning apps where we are trying to predict how much something should cost, or how many people will be at a location, or how often an event will happen. But - it's not enough to just make a prediction: we should also be able to check our accuracy to see how we did.

 **Distribute:** Pass out the **Numerical Accuracy** to each student

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 the slide with the overview of this task - the class has been rescued and can now see how many zombies were actually at Locations A, B, and C. Using this information, students can start to calculate the accuracy of their results.

Teaching Tip

Accuracy and Numerical Data: Accuracy is calculated differently for numerical data than it is for categorical data. This part of the activity is important for building conceptual understanding so when students see these same calculations in AI Lab, they'll have a reference for understanding what it means. Skipping too fast through this section may mean students are confused when they see accuracy in AI Lab with numerical data.

 **Discuss:** Based on the results, what was the accuracy of our model? How many locations did it predict exactly correct?

Discussion Goal: Students should notice that our model is 0% accurate if we were expecting it to be perfectly correct. Guide students to consider whether it's okay for the model to be "close enough" - for example, when weather apps predict the temperature, we let it be okay if they're off by a few degrees as long as they are close enough.

 **Discuss:** What would the accuracy of our model be if it was okay that we were within 5 of the actual value? If we were within 20?

Discussion Goal: Students should notice that if the number is too large (like within 20), then the model appears highly accurate, and if the number is lower than the model appears less accurate.

 **Do This:** Look at the data in the table below and calculate the accuracy of our model using these three different approaches.

Circulate: Check in with students as they calculate the remaining accuracy. An answer key is provided with the correct answers.

Remarks

When making predictions with numerical data, it's rare for the prediction to be an exact match with the actual number. Instead, we can check if the prediction is "close enough" to the actual value and base our accuracy from that. But, deciding what "close enough" is depends on the situation. Tomorrow, we'll see how AI Lab makes predictions with numerical data and how to understand it's accuracy.

 **Discuss:** In one of the rows, the model predicted 0 but the actual value was 2. Is this close enough to count as a correct prediction?

Discussion Goal: This should be a quick discussion, where students will likely say that yes - this is close enough to count as correct.

 **Discuss:** In one of the rows, the model predicted 0 **zombies** but the actual value was 2 **zombies**. Even though this is close enough to be correct, is that okay for the people in this situation?

Discussion Goal: Guide students to remembering that even though the data represents numbers, the situation involves people under specific circumstances. Even though the situation is from a science fiction story, if we put ourselves in the shoes of the people in this city: this means they likely went to a location where they weren't expecting any zombies and instead found two. Even though the numbers were close enough to be correct, this small margin of error had a huge effect on the people involved.

AI Accuracy in the Real World: This situation introduces issues involving inaccurate calculations from machine learning models, such as false positives or false negatives. Students may have experience with this when an important email is marked as spam by an email program, or they ignore a phone call from an important person because they don't recognize the number - both of these are examples of false positives, where a piece of data was incorrectly marked as unimportant.

Not all false positives have the same impact on people in the real world. These resources can be helpful in giving examples of how issues of accuracy can have large impacts for people in the real world:

- **Medical News and Life Sciences** - this article explains how artificial intelligence is routinely used to help identify cancer in the medical field to great success, but the risks of a false-positive identification can be devastating.
- **Wrongfully Accused by an Algorithm** - this article explains how inaccurate facial recognition software was used to wrongfully arrest a man for a crime he didn't commit.

Wrap Up (5 minutes)

Journal

Prompt: *What are other situations where you think machine learning models should be exactly accurate because close-enough isn't okay?*

Discussion Goal: Students may call on examples from the news such as:

- Medical diagnosis, such as cancer detection
- Facial recognition for criminal prosecution



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**.