Handwriting Calculator Website

Brighton Mica, Nathan Swedlund

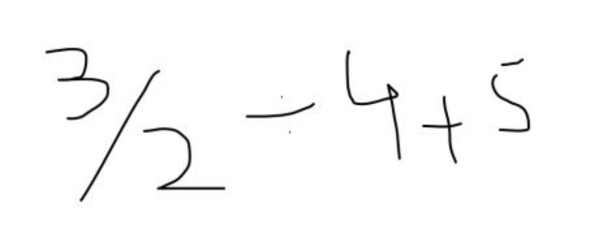
Problem Statement

Research Question

What methods and techniques can we use to identify and accurately classify handwritten numerals and mathematical notation while retaining the relationships between them so that we can reduce such input to a numerical calculation?

What We Are Trying To Do

We are trying to make a classifier for handwritten digits and mathematical notation so that we can reduce handwritten math problems to their numerical answer.



3/2 - 4 + 5 = 2.5

Datasets

Scikit-learn Digits

- Images: 1797
- Features: 64 (8x8 pixels)
- Digits `0-9`

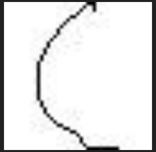
URL: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_digits.html

Handwritten Math Symbols

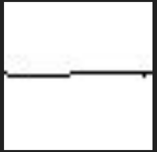
- Images: 100,000+
- Features: 2025 (45x45 pixels)
- `+, -, /, ), (, *`

URL: https://www.kaggle.com/xainano/handwrittenmathsymbols

(

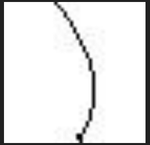


-



+

)



How Much Data Will We Need?

Though it’s hard to say how much data we will need, we hope that our two datasets (digits, and handwritten math symbols) will be sufficient. If they are not, we will either find more datasets, or create our own data.

Pattern Recognition Techniques & Classifiers

Classifiers:

- **GuassianNB**
- **Decision Tree**
- **K-neighbors**
- **Neural Network**

Our approach:

We intend to test the feasibility of all the classifiers above. Though, we would like to use a decision tree if possible. This is because if we were to go to production with this site, we could run a decision tree classifier with much less computation on the backend.

Design Choices & Parameter Optimization

Design Choices:

Segmenting

Parameter Optimization:

Several of the classifiers we intend on using will have parameters we can adjust. We will be using a test/train split to find the optimal values for these parameters. Specifically we will be optimizing for accuracy. Even though time efficiency is a concern, we would rather the answer to be correct and slow than fast and incorrect.

Data Analysis & Output

Data Analysis:

Most of the data analysis we will be doing is parameter optimization. For each classifier we test, we will make graphs showing the relationship between accuracy and its parameters (if it has any).

We will also have a conclusion section at the end of the jupyter notebook justifying our final choice of classifier.

Output:

The output for our classifiers will be routed directly to the web site so it will be visible to the user.

Presentation of Results

Web Site (local or hosted)

The website will be our final product where one can use our classifier. Users will be able to write out math problems within a text field using their mouse, stylus, etc., and the answer will be displayed.

The website is a frontend presentation of our results as it does not display any information concerning our methodology.

Research Paper

Our reasonings and analytical findings will be concluded in an ACM-style paper. This paper will be a summation of the work we've done covering points such as

- What methods performed well?
- What methods performed unfavorably?
- What optimizations were made?
- Is there any future work?

Link To Repo: https://github.com/NathanSwedlund/HandwritingCalculator

Tasks per Member

Both of us intend to work of every aspect of this project. As documentation, we will include a to-do list on the repo with our names by what each person did. Also, the commit log on git will show who did what.

- Create Classifiers
  - Neural Net
  - Decision Tree
  - K-Neighbors
  - Gaussian Naive Bayes
- Parsing image for individual characters

- Function to solve math equation given an input string
- Optimizations
- Optional
  - Deploying to Heroku
  - Real-Time (on pen-up update screen)
  - UI Improvements

Timeline Projections

Already done:

- UI for reading in user’s writing
- Code to convert handwriting to an image.
- The Django Server

Week 1 (Apr 12 - 18):

- Basic classifiers
- System to parse each character from the input image

Week 2 (Apr 19 - 26):

- Optimization to classifiers
- Finalization to Jupyter Notebook
- Math solving function

Week 3 (Apr 27 - May 3):

- Optimizations
- Pushing project to Heroku

Links:

Image used in slide 2:

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.addictivetips.com%2Fwindows-tips%2Finkulator-for-windows-8-rt-solves-handwritten-math-expressions%2F&psig=AOvVaw09lQAOE5lkmngmoHPAoc5V&ust=1586553946512000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCIjyvJik3OgCFQAAAAAdAAAAABAE

Datasets:

https://scikit-learn.org/stable/modules/generated/ sklearn.datasets.load_digits.html

https://scikit-learn.org/stable/modules/generated/ sklearn.datasets.load_digits.html

Github Repository:

https://github.com/NathanSwedlund/HandwritingCalculator