K-12 Data Science Curriculum

ELECTRICAL & COMPUTER ENGINEERING

Sponsor: Microsoft
Mentor: Shana Matthews
Faculty Advisor: Payman Arabshahi
Students: Abhishek Sangameswaran, Kailing Shen

Microsoft

Development of a hands-on Data Science Curriculum which aims to make K-12 learners data literate as well as introduce them to Data Science topics such as Data Visualization, Statistics, and Machine Learning.

Introduction

There is a significant gap between data literacy and data science – the skills required to succeed in tomorrow's workforce will require domain-specific knowledge and coding skills alike. Data Science / Machine Learning / AI are significant areas of innovation, but we lack the learning paths and hands-on inspiration necessary to motivate young learners.

Requirements

After conducting student interviews and reflecting carefully on the goals of the curriculum, these five requirements were identified:

1. Short video-based lessons.
2. Teach using visuals / help build intuition.
3. Promote critical thinking.
4. Hands-on exercises and projects.
5. Allow some creative freedom.

Curriculum Outline

Chapter I: Fundamentals of Data
- Introduction to data, data collection, and data interpretation.

Chapter II: Data Visualization
- Introduction to basic data visualizations and Excel.

Chapter III: Statistics
- Introduction to descriptive & inferential Statistics.

Chapter IV: Machine Learning
- Introduction to Machine Learning algorithms and python.

Chapter V: Why Pursue Data Science?
- Motivation to pursue Data Science, overview of educational requirements, and career prospects.

Curriculum Design & Implementation

Each unit / chapter has three components: Lessons, Exercises, and Project. The lessons introduce the topic, the exercises present guided problems for students to check their understanding of concepts , and the projects present problems that are less guided and require students to choose and apply concepts they have learnt. Additionally, each unit follows these guidelines:

1. Introduce topics with examples.
2. Teach the underlying mathematics behind each concept.
3. Use copious amounts of visuals to explain the concept.

Presented below is a snapshot of the Machine-learning unit subchapter on the Perceptron algorithm:

1. Lesson:

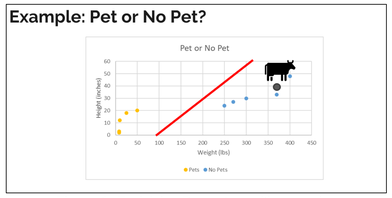

Figure 1: Explaining Perceptron with example

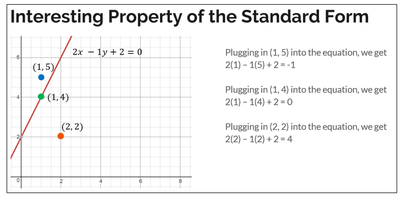

Figure 2: Explaining the math behind Perceptron

2. Exercise:

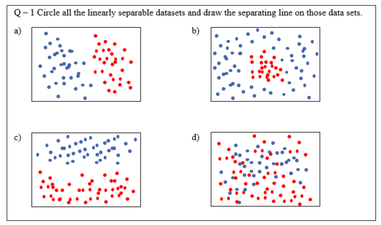

Figure 3: Identifying linearly separable data

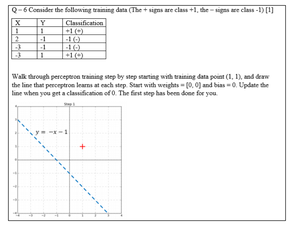

Figure 4: Perceptron algorithm walkthrough

3. Project:

```
def vector_add(a, b):
  result = []
  for i in range(len(a)):
    # Your code here

  return result

def vector_scalar_multiply(scalar, vector):
  result = []
  for i in range(len(a)):
    # Your code here

  return result

def sign(floating_point_number):
  # Your code here
```

Figure 5: Perceptron support functions

```
def perceptron_train(training_data):

  weights = [0 for i in range(len(training_data[0]) - 1)]
  bias = 0
  epochs = 100

  for e in range(epochs):
    for data in training_data:
        training_data_element = data[1:]
        actual_classification = data[0]
        guess = sign(dot_product(weights, training_data_element) + bias)

        if(guess != actual_classification):
            # Your code here
  return weights, bias
```

Figure 6: Perceptron train function

Future Work

To begin testing this curriculum with students, the first step is to film a voiceover of the lessons. Once the video lessons are complete, there are five areas of future work that will help fully mature this curriculum:

1. Test the curriculum with a sizable student population.
2. Refine content based on student feedback.
3. Host the curriculum online.
4. Automate checking of exercises where possible.
5. Setup online forum where students can ask questions.

Conclusion

The goal of this project was to design a hands-on curriculum that builds a solid foundation of the Data Science areas of knowledge in young learners, and inspires them to consider data science as a career. The curriculum accomplishes this by introducing students to the Fundamentals of Data, Data Visualization, Statistics, Machine Learning, and Python. To inspire students to pursue Data Science, the curriculum includes a subchapter dedicated to contemporary problems that have been solved with Data Science.

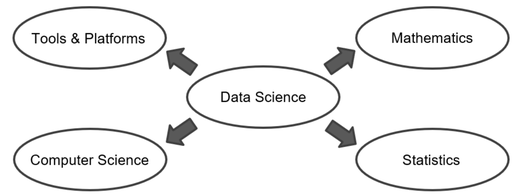

Figure 7: Data Science areas of knowledge

References

[1] Machine Learning Introduction by Noah Smith, https://courses.cs.washington.edu/courses/cse446/17au/intro.pdf
[2] Introduction to Machine Learning by Eric Grimson, https://www.youtube.com/watch?v=h0e2HAPTGF4
[3] Perceptron by Noah Smith, https://courses.cs.washington.edu/courses/cse446/17au/perceptron.pdf
[4] Perceptron Exercise by Noah Smith, https://courses.cs.washington.edu/courses/cse446/17au/A2.pdf