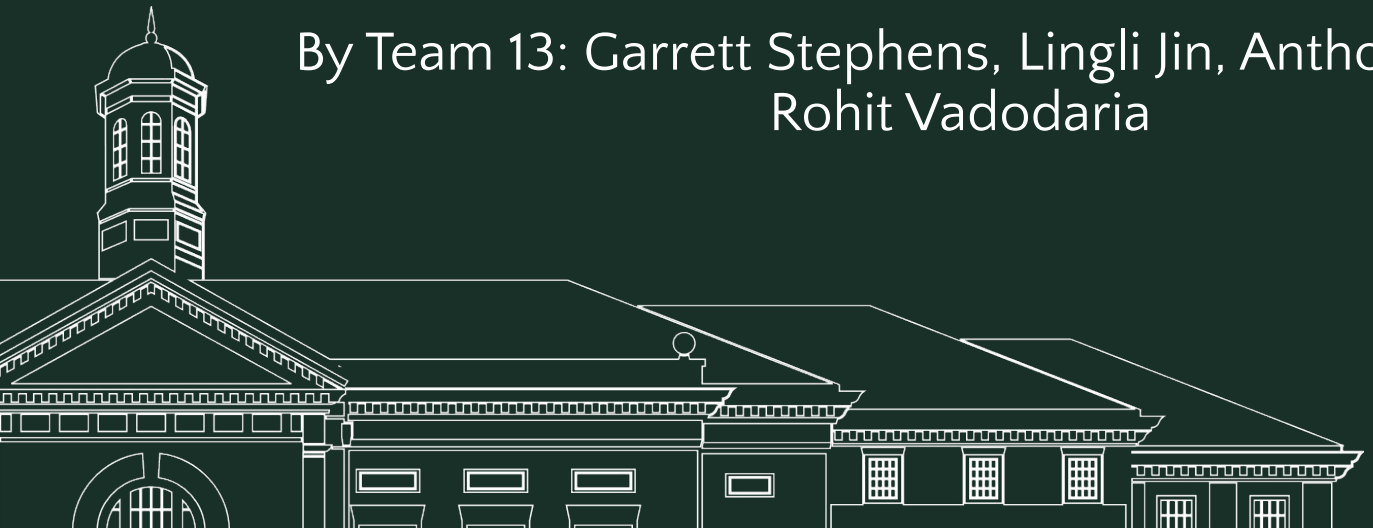
Raymond A. Mason
School of Business
WILLIAM & MARY

Stochastic Gradient Descent

By Team 13: Garrett Stephens, Lingli Jin, Anthony Morris, Rohit Vadodaria



Agenda

- What is stochastic gradient descent?
- Why would you use it?
- What are some limitations?
- What does it look like in code?

What is Stochastic Gradient Descent?

- A machine learning optimization algorithm that is used to find the best fit of actual and predicted values
- Uses sample “batches” of data until it iterates through the whole dataset
- Supports multi-class classification and regression problems
- Common applications found in text classification and natural language processing

Why Use SGD?

- Helpful for large-scale machine learning problems
- Can scale to datasets with over 100000 training samples and 100000 training features
- Results in a very efficient approach to training machine learning algorithms
- Lots of opportunities for code tuning

What Are Limitations of SGD?

- **Sensitivity to Feature Scaling**
- **Convergence rate**
- **Local Minima**
- **Hyperparameters**
- **Oscillation**
- **Noise Tolerance**

Overall, SGD has proven to be effective in many applications, but it is important to understand its limitations and to use other optimization algorithms when necessary.

What Does SGD Look Like in Code?

SGDdemo.ipynb - Colaboratory (google.com)