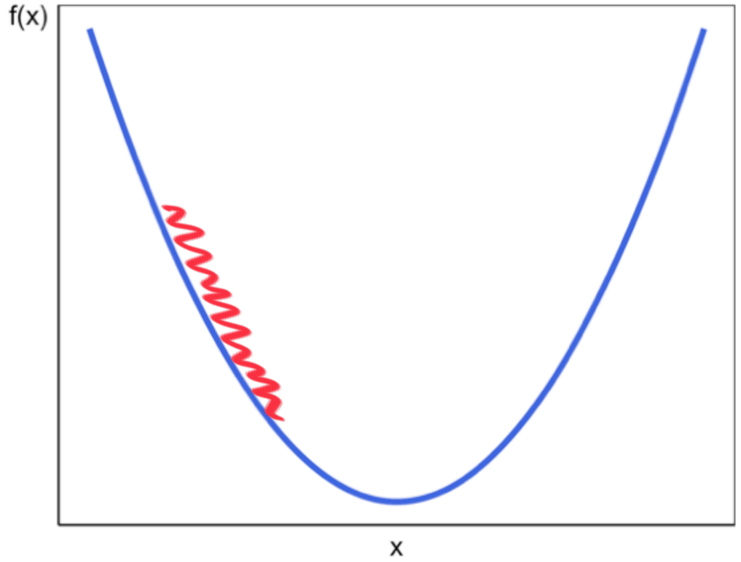
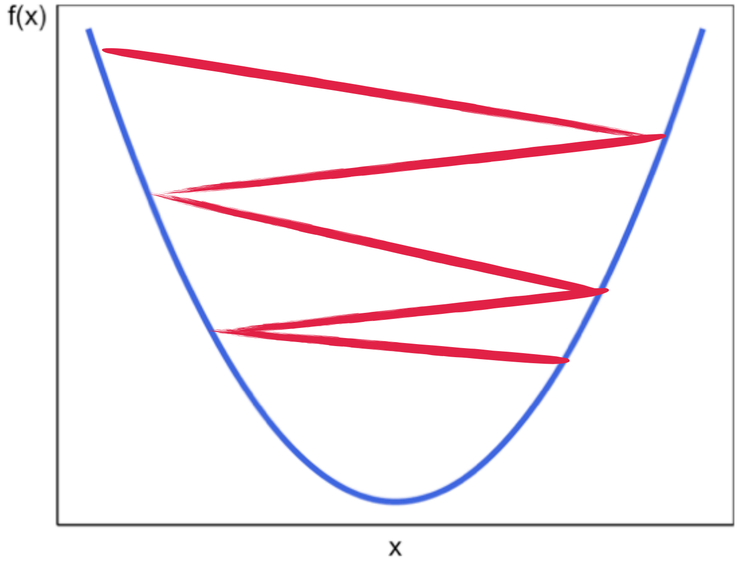
Choosing a Learning Rate: Convexity

- Under ideal conditions, gradient descent iteratively approximates and converges to the optimum
- For a constant learning rate λ
  - if λ is too small, it takes too many iterations to reach the optimum
  - if λ is too large, algorithm may 'bounce' around the optimum and never get close





  - Better to treat learning rate as a variable, that is let the value depend on gradient
  - around optimum λ is small, and far from optimum λ is larger

This Week's Practical