Exercise 1 - Introduction

November 28, 2019

1   Linear Algebra Refresher

a) Matrix Operations (1 point)

A fellow student suggests that matrix addition and multiplication are very similar to scalar addition and multiplication, i. e. commutative, associative and distributive. Is this a correct statement? Prove it mathematically or disprove it by providing at least one counter example per property (commutativity, associativity, distributivity).

b) Matrix Inverse (1 point)

What is a matrix inverse? How can you build the inverse of a non-square matrix? You would like to invert a matrix $M \in \mathbb{R}^{2 \times 3}$ - write down the equation for computing it and specify the dimensionality of the matrices after each single operation (e.g. multiplication, inverse).

c) Eigenvectors and Eigenvalues (1 point)

Explain what eigenvectors and eigenvalues of a matrix M are. Why are they relevant in machine learning?

2 Statistics Refresher

a) Terminology (1 point)

What is a random variable? What is a probability density function (PDF)? What is a probability mass function (PMF)? What do a PDF and a PMF tell us about a random variable?

Applied Machine Learning Fundamentals

b) Expectation and Variance (1 point)

State the general definition of expectation and variance for the probability density $f : \Omega \to \mathbb{R}$ of a continuous random variable. What do expectation and variance express?

3 Optimization

a) Numerical Optimization - Gradient Descent (5 points)

Implement a simple gradient descent algorithm for finding a minimum of the Rosenbrock function with $n = 2$ using Python and NumPy:

$$f(x) = \sum_{i=0}^{n-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right]$$

Submit your code and a plot of the learning curve for the best run of your gradient descent implementation. Which learning rate worked best? (Hint: You need to find the first derivative(s) of $f(x)$ for $n = 2$ and iteratively evaluate them during gradient descent. Automatic differentiation tools are not allowed for this exercise. Choose a random starting point for the parameters, for example $x_i \in [-2, 2]$.)

4