Tutorial 4 Problems

CSC311, Fall 2021

1 Gradient Descent Intuition

Suppose we are trying to optimize the loss function $f(x) = \frac{1}{2} x^T A x$, where $x \in \mathbb{R}^2$

1. Let $A = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ and $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

   What are the first two iterates of gradient descent, with a learning rate $\eta = 0.1$?

2. For which learning rates will gradient descent converge? The convergence speed is determined by how the error decreases in the "slowest" direction. What learning rate leads to the fastest convergence?

3. Suppose we choose the optimal learning rate. How many steps of gradient descent does it take for both components to be less than 1e-3 (0.001)?

4. Repeat the previous two parts with $A = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$.

2 Sum of Convex Functions

Prove that the sum of two convex functions is convex.