DSE 220: Machine learning

Worksheet 7 — Unconstrained optimization

1. We are given a set of data points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, and we want to find a single point $z \in \mathbb{R}^d$ that minimizes the loss function
$$L(z) = \sum_{i=1}^n \|x^{(i)} - z\|^2.$$
Use calculus to determine z, in terms of the $x^{(i)}$.

2. Given a set of data points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, we want to find the vector $w \in \mathbb{R}^d$ that minimizes this loss function:
$$L(w) = \sum_{i=1}^n (w \cdot x^{(i)}) + \frac{1}{2} c \|w\|^2.$$
Here $c > 0$ is some constant.
   (a) What is $\nabla L(w)$?
   (b) What value of w minimizes $L(w)$?

3. Consider the following loss function on vectors $w \in \mathbb{R}^4$:
$$L(w) = w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4.$$
   (a) What is $\nabla L(w)$?
   (b) Suppose we use gradient descent to minimize this function, and that the current estimate is $w = (0, 0, 0, 0)$. If the step size is η, what is the next estimate?
   (c) What is the minimum value of $L(w)$?
   (d) Is there is a unique solution w at which this minimum is realized?

4. Consider the loss function for ridge regression (ignoring the intercept term):
$$L(w) = \sum_{i=1}^n (y^{(i)} - w \cdot x^{(i)})^2 \; + \; \lambda \|w\|^2$$
where $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)}) \in \mathbb{R}^d \times \mathbb{R}$ are the data points and $w \in \mathbb{R}^d$. There is a closed-form equation for the optimal w (as we saw in class), but suppose that we decide instead to minimize the function using local search.
   (a) What is $\nabla L(w)$?
   (b) Write down the update step for gradient descent.
   (c) Write down a stochastic gradient descent algorithm.