CS 217: Artificial Intelligence and Machine Learning **Jan 10, 2024**

Lecture 1: The Basics of Optimization

Lecturer: Swaprava Nath *Scribes: Aditya Singh, Dion Reji, Shreyas Katdare, Brian Mackwan*

Disclaimer: *These notes aggregate content from several texts and have not been subjected to the usual scrutiny deserved by formal publications. If you find errors, please bring to the notice of the Instructor.*

In this lecture, we discuss some basic optimization problems.

	Variable	Solution space	Solution complexity
Continuous optimization	take any real value (continuous)	infinite	polynomial in the size of the problem
Discrete optimization	discrete	finite	exponential (e.g., knapsack)

1.1 What is optimization?

An optimization problem involves finding the minimum value attained by a function subject to some constraints, i.e.,

$$\min_{x \in \mathscr{C}} f(x),$$

where $f(x)$ is the objective function and $\mathscr{C}$ is the constraint set.

Example. Minimize $(x-2)^2$ with the constraint that $x \in [0,1] \cup [4,7]$.

Here, the objective function is $f(x) = (x-2)^2$ and the constraint set is $\mathscr{C} = [0,1] \cup [4,7]$. This is a one-variable function, and we can easily see from the plot in figure 1.1 that $x^* = 1$ is the optimal value of x.

Example. We want to optimally place a warehouse, so that the sum of the Euclidean distances between the warehouse and the cities is minimized.

Let the cities be located at $\mathbf{y}_1, \ldots, \mathbf{y}_m$ and the warehouse at $\mathbf{x}$. We want to find the following:

$$\min_{\mathbf{x} \in \mathscr{C}} \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{y}_i\|_2,$$

where $\mathscr{C}$ is the set of all points where we want our warehouse to be, and $\|\cdot\|_2$ represents the L^2 norm, defined by

$$\|\mathbf{x}\| := \sqrt{x_1^2 + x_2^2},$$

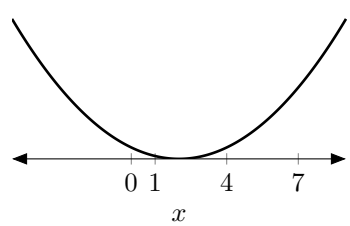

Figure 1.1: A plot of $f(x) = (x-2)^2$.

where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Example (Image de-blurring). We consider grayscale images of size $m \times n$, where each pixel has an intensity value in $[0, 1]$. The input image is $\mathbf{y} = [y_{i,j}]^{m \times n}$ and the desired output is $\mathbf{x}$.

$$\min_{\mathbf{x} \in [0,1]^{m \times n}} \left(\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|y_{i,j} - (k * \mathbf{x})_{i,j}\| + \lambda \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} ((\mathbf{x}_{i,j} - \mathbf{x}_{i,j+1})^2 + (\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j})^2) \right)$$

Here, λ and k are the hyperparameters, which are determined by experiment. The optimal value of λ and k depend on the image. For example, medical images may require different λ and k as compared to images of trees.

Example (Machine learning; curve-fitting). We have inputs (x_i, y_i), where $i \in [n]$. We want to find

$$\min_{\Theta} \sum_{i=1}^{n} \ell(h_\theta(x_i), y_i),$$

where $\ell(\cdot, \cdot)$ is a loss function, $h_\Theta(x) = w_0 + w_1 x + w_2 x^2$ is the hypothesis (so $h_\theta(x_i)$ is the hypothesized point), and $\Theta = (w_0, w_1, w_2)$. An example of a loss function is

$$\ell(y_1, y_2) = (y_1 - y_2)^2.$$