ELEC 378 – Spring 2023
Homework 4

Due: Friday February 10, 5PM

1 The Gradient of a Linear Functional

Derive an expression for the gradient of each of the following loss functions with respect to $\mathbf{w} \in \mathbb{R}^p$:

a) $\mathcal{L}(\mathbf{w}) = \langle \mathbf{a}, \mathbf{w} \rangle$, $\mathbf{a} \in \mathbb{R}^p$

b) $\mathcal{L}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{w} \rangle$

2 Gradient Descent vs Stochastic Gradient Descent

Let $\mathcal{L}(\mathbf{w}) = \sum_{i=1}^{3} \mathcal{L}_i(\mathbf{w})$ with $\mathbf{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^\intercal$ and

$$\begin{aligned}
\mathcal{L}_1(\mathbf{w}) &= 2w_1^2 + w_1 w_2 - 4w_2^2 \\
\mathcal{L}_2(\mathbf{w}) &= 3w_1^2 + 4w_1 w_2 + 5w_2^2 \\
\mathcal{L}_3(\mathbf{w}) &= -w_1^2 - 4w_1 w_2 + 3w_2^2
\end{aligned}$$

Derive both gradient descent and stochastic gradient descent algorithms for finding the minimizer $\mathbf{w}^\star = \underset{\mathbf{w} \in \mathbb{R}^2}{\operatorname{argmin}}\ \mathcal{L}(\mathbf{w})$, and implement them both from scratch. Do you notice any differences between the two algorithms?

3 Color Quantization

To reduce the size of an RGB image file, we can decrease the maximum possible number of distinctly colored pixels. One method to accomplish this is to choose K distinct pixel colors and replace each pixel in the image with the closest of the K chosen pixel colors via K-means clustering. Instead of storing each pixel using 24 bits (8 bits each for R, G, B channels), we can then store each pixel using $\log_2(K)$ bits.

a) While we often model each pixel in an RGB image as a vector in $\mathbb{R}^3$, the RGB coordinates are actually each quantized to 8 bits. With this in mind, what is the maximum possible number of distinctly colored pixels in an arbitrary RGB image?

b) To perform K-means clustering, one must first form the data matrix $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \dots & \mathbf{x}_n\end{bmatrix}^\intercal$. What should the feature vectors $\{\mathbf{x}_i \in \mathbb{R}^p\}_{i=1}^n$ represent in this scheme? What is p? What is n?

p = 3, n = # of pixels, feature vectors are individual pixels

c) Form the data matrix for the image `objection.png` and apply PCA to reduce the dimensionality to 2. View the scatter plot of the dimensionality-reduced data. Is it clear how the feature vectors should be clustered?

d) Implement K-means clustering from scratch to quantize `objection.png` using $K = 64 = 2^6$, so that each pixel can be stored using 6 bits instead of 24. Include in your writeup both the quantized image and your PCA scatterplot colored with the resulting quantization to see the Voronoi tiling. **Note: It is acceptable if your implementation generates fewer than 64 clusters provided you started with 64.**

4 Dr. Data Science II

Last week we were able to use PCA to observe a cluster of melanoma patients' gene expression data in 2D that allowed us to identify genes that are important in making a melanoma diagnosis. However, this cluster contained two errant breast cancer patients, and one melanoma patient was placed outside this cluster. Use the `sklearn` implementation of K-means to find a more robust clustering of melanoma patients in higher dimensions, that is, still using PCA to reduce the dimensionality of the data but retaining more than just two principal components. What is the smallest number of retained principal components that allows for a more robust clustering?

Submission Instructions

Every student must submit their work in PDF format, providing intermediate and final results as well as any necessary code. Submit your homework on Gradescope.

Collaboration Policy

Collaboration both inside and outside class is encouraged. You may talk to other students for general ideas and concepts, but individual write-ups must be done independently.

Plagiarism

Plagiarism of any form will not be tolerated. You are expected to credit all sources explicitly.