Discussion 13

1. **Sphere Enclosure**

   For $i = 1, \dots, m$, let B_i be a ball in $\mathbb{R}^n$ with center $\vec{x}_i$, and radius $\rho_i \geq 0$. We wish to find a ball B of minimum radius that contains all the B_i for $i = 1, \dots, m$. Cast this problem as an SOCP.

2. LASSO vs. Ridge

Consider the data set $\{(\vec{x}^{(i)}, y^{(i)})\}_{i=1,\ldots,n}$ of samples $\vec{x}^{(i)} \in \mathbb{R}^d$ and values $y^{(i)} \in \mathbb{R}$. Define $X = \begin{bmatrix} \vec{x}^{(1)} & \ldots & \vec{x}^{(n)} \end{bmatrix}^\top \in \mathbb{R}^{n \times d}$ and $\vec{y} = \begin{bmatrix} y^{(1)} & \ldots & y^{(n)} \end{bmatrix}^\top \in \mathbb{R}^n$, i.e., X is the $n \times d$ matrix whose i-th column is $(\vec{x}^{(i)})^\top$, for each $i \in \{1, \cdots, n\}$, and $\vec{y}$ is the n-dimensional column vector whose i-th component is y_i, for each $i \in \{1, \cdots, n\}$.

For the sake of simplicity, assume that the data has been centered and whitened so that each feature has mean 0 and variance 1 and the features are uncorrelated, i.e. $X^\top X = nI_{d\times d}$, where $I_{d\times d}$ denotes the $d \times d$ identity matrix. Consider the linear least squares regression with regularization in the ℓ_1-norm, also known as LASSO:

$$\vec{w}^\star = \underset{\vec{w} \in \mathbb{R}^d}{\operatorname{argmin}} \; \|X\vec{w} - \vec{y}\|_2^2 + \lambda \|\vec{w}\|_1 . \tag{1}$$

This problem will compare ℓ_1-regularization with ℓ_2-regularization (ridge regression) to understand their similarities and differences, by looking at the elements of $\vec{w}^\star$ in the solution to each problem.

(a) First, decompose this optimization problem into d univariate optimization problems over each element of $\vec{w}$. *Hint*: Let $\vec{x}_j \in \mathbb{R}^n$ denote the j-th column of X, so that $X = \begin{bmatrix} \vec{x}_1 & \ldots & \vec{x}_d \end{bmatrix}$ and recall that $X^\top X = nI_{d\times d}$.

(b) Prove that for any $i \in \{1, \cdots, d\}$, if $\vec{y}^\top \vec{x}_i > \frac{1}{2}\lambda$ then $w_i^\star > 0$. Find $w_i^\star$ in that case.

(c) Prove that for any $i \in \{1, \cdots, d\}$, if $\vec{y}^\top \vec{x}_i < -\frac{1}{2}\lambda$ then $w_i^\star < 0$. Find $w_i^\star$ in that case.

(d) Prove that for any $i \in \{1, \cdots, d\}$, if $|\vec{y}^\top \vec{x}_i| < \frac{1}{2}\lambda$ then $w_i^\star = 0$.

(e) Now consider the case of ridge regression, which uses the the ℓ_2 regularization $\lambda \|\vec{w}\|_2^2$.

$$\vec{w}^\star = \underset{\vec{w} \in \mathbb{R}^d}{\operatorname{argmin}} \; \|X\vec{w} - \vec{y}\|_2^2 + \lambda \|\vec{w}\|_2^2 . \tag{2}$$

Write down the new condition for $\vec{w}_i^\star$ to be 0. How does this differ from the condition obtained in part (4) and what does this suggest about LASSO?

