THU-70250403, Convex Optimization (Fall 2021) **Homework: 7**

Classification and Parameter Estimation

Lecturer: Li Li `li-li@tsinghua.edu.cn`

Student:

Problem 1

Please prove that the loss function of a deep neural network is generally non-convex.

Hint: please consider the following simplest deep neural network as an example, which has one hidden layer. Given m smples $(\boldsymbol{x}_i, y_i)$, $\boldsymbol{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$ $(i = 1, \ldots, m)$, the loss function is:

$$\min_{\boldsymbol{w}_1, \boldsymbol{w}_2} \quad J(\boldsymbol{w}_1, \boldsymbol{w}_2) = \frac{1}{m} \sum_{i=1}^{m} \left[y_i - \left(\sum_{j=1}^{h} w_{2,j} \delta \left(\boldsymbol{w}_{1,j}^T \boldsymbol{x}_i + b_{1,j} \right) + b_2 \right) \right]^2 \tag{1}$$

where $\boldsymbol{w}_{1,j} \in \mathbb{R}^n$, $w_{2,j} \in \mathbb{R}$ $(j = 1, \ldots, h)$ are the weights to be optimized. $b_{1,j} \in \mathbb{R}$ $(j = 1, \ldots, h)$, $b_2 \in \mathbb{R}$ are pre-chosen biases. The transfer function $\delta(\cdot) : \mathbb{R} \to \mathbb{R}$ could be $\delta(z) = \max\{0, z\}, z \in \mathbb{R}$.

Problem 2

Please derive the dual problem of the following primal problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & |A\boldsymbol{x} - \boldsymbol{b}|_2^2 & (2) \\ \text{s.t.} \quad & \boldsymbol{x}^T \mathbf{1} \leq 1 & (3) \\ & \boldsymbol{x}\boldsymbol{x}^T \leq \lambda I & (4) \end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^+$, $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = m$.

Problem 3

(This problem is Optional)

Please explain how to handle linear measurement with noise, where the noise follows a mixture of K univariate normal distributions [1].

References

[1] W. Yao, Y. Wei, C. Yu, "Robust mixture regression using the t-distribution," *Computational Statistics and Data Analysis*, vol. 71, pp. 116-127, 2014.