CS2601 Linear and Convex Optimization

Homework 1

Due: 2022.9.26

1. For $\boldsymbol{x} = (x_1, x_2)^T \in \mathbb{R}^2$, determine whether the following functions have a global minimum and show your arguments. You do **NOT** have to find the global minima.

(a). $f(\boldsymbol{x}) = 2x_1^2 + 2x_1x_2 + 3x_2^2 - x_1 - 2x_2$. Hint: Show $f(\boldsymbol{x})$ is coercive.

(b). $f(\boldsymbol{x}) = x_1^2 + 2x_1x_2 + 3x_2^2 - x_1 - 2x_2$. Hint: Show $f(\boldsymbol{x}) \geq \frac{1}{2}\|\boldsymbol{x}\|^2 - 1$.

(c). $f(\boldsymbol{x}) = x_1^2 + 2x_1x_2 + x_2^2 - x_1 - 2x_2$. Hint: Consider its restriction to a line.

2. Find the gradient of the following functions,

(a). $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|^2$

(b). $f(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|^2$, where $\boldsymbol{X}, \boldsymbol{y}, \lambda$ are known constants.

3. **Logistic regression.** Recall the objective function of logistic regression is the following negative log likelihood,

$$f(\boldsymbol{w}) = \sum_{i=1}^{m} \log(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}),$$

where $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^n \times \{-1, +1\}$ is the i-th data point. We have absorbed the bias term b into $\boldsymbol{w}$ by appending an extra 1 to each $\boldsymbol{x}_i$.

(a). Suppose the dataset $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_m, y_m)\}$ is strictly linearly separable in the sense that there exists a $\boldsymbol{w}_0$ such that

$$y_i \boldsymbol{x}_i^T \boldsymbol{w}_0 > 0, \quad \forall i = 1, 2, \ldots, m.$$

Does f have a global minimum in this case? Explain your answer.

(b). Suppose the dataset is not linearly separable in the sense that for any $\boldsymbol{w}$, there exists an $i_0 = 1, 2, \ldots, m$ such that

$$y_{i_0} \boldsymbol{x}_{i_0}^T \boldsymbol{w} < 0.$$

Show that f has a global minimum by completing the following steps.

i) Show

$$f(\boldsymbol{w}) \geq h(\boldsymbol{w})$$

where

$$h(\boldsymbol{w}) = \max_{1 \le i \le m} -y_i \boldsymbol{x}_i^T \boldsymbol{w}.$$

ii) Let $S = \{\boldsymbol{w} : \|\boldsymbol{w}\| = 1\}$ be the unit sphere. Show that $h(\boldsymbol{w})$ has a global minimum $\boldsymbol{w}_0$ on S and $C \triangleq h(\boldsymbol{w}_0) > 0$. You can assume the fact that h is continuous, which can be proved by induction and the identity $\max\{a, b\} = \frac{a+b+|a-b|}{2}$.

iii) Show

$$h(\boldsymbol{w}) \ge C\|\boldsymbol{w}\|, \quad \forall \boldsymbol{w}$$

iv) Show f has a global minimum.

(c). Find $\nabla f(\boldsymbol{w})$.

(d). Now we add a regularization term to the objective function, i.e. we consider

$$\tilde{f}(\boldsymbol{w}) = \sum_{i=1}^{m} \log(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}) + \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2,$$

where $\lambda > 0$. Does $\tilde{f}(\boldsymbol{w})$ have a global minimum? Does your answer depend on whether the dataset is linearly separable?