CS2601 Linear and Convex Optimization

Homework 1

Due: 2021.9.30

1. Let $f(\boldsymbol{x}) = 2x_1^2 + x_1x_2 + x_2^2 - 3x_1 - 5x_2$.

(a). Is $f(x)$ coercive? Hint: use $x_1x_2 \geq -(x_1^2 + x_2^2)/2$.

(b). Find the minimum and maximum of $f(\boldsymbol{x})$ over $\mathbb{R}^2$ if they exist.

2. 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

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

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

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

iv) Show f has a global minimum.

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

3. **Taylor expansion.** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

(a). Show that the following first-order Taylor expansion with Lagrange remainder,

$$f(\boldsymbol{x}+\boldsymbol{d}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}+t\boldsymbol{d})\boldsymbol{d}$$

for some $t \in (0,1)$. You can assume the same expansion for univariate functions.

(b). Recall the integral of a vector valued function $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^n$ is defined component-wise, i.e. for $\boldsymbol{g}(t) = (g_1(t), g_2(t), \ldots, g_n(t))^T$,

$$\int \boldsymbol{g}(t)dt = \begin{pmatrix} \int g_1(t)dt \\ \int g_2(t)dt \\ \vdots \\ \int g_n(t)dt \end{pmatrix}$$

Show

$$\nabla f(\boldsymbol{x}+\boldsymbol{d}) = \nabla f(\boldsymbol{x}) + \int_0^1 \nabla^2 f(\boldsymbol{x}+t\boldsymbol{d})\boldsymbol{d}\; dt$$

Hint: Apply the Newton-Leibniz formula to $\boldsymbol{g}(t) = \nabla f(\boldsymbol{x}+t\boldsymbol{d})$.

4. Are the following matrices positive definite, positive semidefinite, negative definite, negative semidefinite, or indefinite?

$$A = \begin{pmatrix} 6 & 2 & 0 \\ 2 & 5 & -2 \\ 0 & -2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & -3 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$