Homework I

2022.9.20

1、(a) Proof: Notice that $f(\vec{x}) = (x_1+x_2)^2 + (x_1-\frac{1}{2})^2 + 2(x_2-\frac{1}{2})^2 - \frac{3}{4}$

Therefore, $\lim_{\|\vec{x}\|\to\infty} f(\vec{x}) = +\infty$ (2 square terms won't be both finite)

$f(x)$ is continuous and coercive, so it has a global minimum.

(b) Proof: $f(\vec{x}) = (\frac{2}{3}x_1 + \frac{3}{2}x_2)^2 + \frac{1}{18}(x_1-9)^2 + \frac{1}{4}(x_2-4)^2 - \frac{17}{2}$

Similarly, $f(x)$ is continuous and coercive, so it has a global minimum.

(c) Proof: Suppose $x_2 = kx_1$, $k\in\mathbb{R}$

Then $f(\vec{x}) = (k+1)^2 x_1^2 - (2k+1)x_1$

Let $k=-1$, Then $f(\vec{x}) = x_1$, so $\lim_{\substack{x_1\to-\infty \\ x_2=-x_1}} f(\vec{x}) = -\infty$

Therefore, $f(x)$ doesn't have a global minimum.

2、(a) $f(x) = \frac{1}{2}x^T x$

$\nabla f(x) = [f'(x)]^T = [\frac{1}{2}\cdot 2x^T I]^T = (x^T)^T = x$

(b) $f(\omega) = \frac{1}{2}\|X\omega - y\|^2 + \frac{\lambda}{2}\|\omega\|^2$

$\nabla f(\omega) = \nabla\left(\frac{1}{2}\|X\omega - y\|^2\right) + \lambda\nabla\left(\frac{1}{2}\|\omega\|^2\right)$

$= X^T\left[(X\omega-y)^T\right]^T + \lambda\omega$

$= X^T(X\omega - y) + \lambda\omega$

$= (X^TX + \lambda E)\omega - X^Ty$

3、(a) Suppose ω_0 separates entire dataset, we have

$y_i x_i^T \omega_0 > 0 \quad , \forall i = 1, 2, \cdots, m$

Now consider $\omega' = t\omega_0$ $(t>0)$,

$y_i x_i^T \omega' = t(y_i x_i^T \omega_0) > 0 \quad , \forall i = 1, 2, \cdots, m$

Then $\lim_{t\to+\infty} f(t\omega_0) = \lim_{t\to+\infty} \sum_{i=1}^{m} \log(1 + e^{-y_i x_i^T \omega'})$

$= \sum_{i=1}^{m} \lim_{t\to+\infty} \log(1 + e^{-t(y_i x_i^T \omega)})$

$= \sum_{i=1}^{m} \log(1+0) = 0$

But $f(\omega) > 0$, $\forall \omega \in \mathbb{R}^n$, which means 0 is not feasible

Therefore, f doesn't have a global minimum.

(b) i. It holds that

$$\log(1+e^{-y_i x_i^T \omega}) > \log(e^{-y_i x_i^T \omega}) = -y_i x_i^T \omega$$

Then

$$\begin{aligned} f(\omega) &= \sum_{i=1}^{m} \log(1+e^{-y_i x_i^T \omega}) \\ &\geqslant \sum_{i=1}^{m} e^{-y_i x_i^T \omega} \\ &\geqslant \max_{1\le i \le m} e^{-y_i x_i^T \omega} \\ &= h(\omega) \end{aligned}$$

ii. Notice: S is compact and h is continuous, by Extreme Value Theorem, $h(\omega)$ has a global minimum $\omega_0 \in S$.

At the same time, for ω_0, there exists an $i_0 = 1, 2, \cdots, m$ s.t.

$$y_{i_0} x_{i_0}^T \omega_0 < 0$$

Then

$$h(\omega_0) = \max_{1\le i \le m} -y_i x_i^T \omega_0 \geqslant -y_{i_0} x_{i_0}^T \omega_0 > 0$$

Therefore, $C \triangleq h(\omega_0) > 0$.

iii. Using homogeneity, let $\omega' = \frac{\omega}{\|\omega\|}$, and $\|\omega'\| = 1$, which means $\omega' \in S$

By conclusion of ii,

$$h(\omega') \geqslant \min_{\omega \in S} h(\omega) = C$$

Then

$$h\left(\frac{\omega}{\|\omega\|}\right) = \frac{1}{\|\omega\|} h(\omega) \geqslant C$$

Therefore, $h(\omega) \geqslant \|\omega\| \cdot C$

iv. We have known that

$$f(\omega) \geqslant h(\omega) \geqslant \|\omega\| \cdot C \quad (C > 0)$$

Then

$$\lim_{\|x\| \to \infty} f(\omega) = +\infty$$

So f is continuous and coersive, it has a global minimum.

(c)

$$\begin{aligned} \nabla f(\omega) &= \sum_{i=1}^{m} \nabla \log(1+e^{-y_i x_i^T \omega}) \\ &= \sum_{i=1}^{m} \left[\frac{1}{1+e^{-y_i x_i^T \omega}} \cdot (-y_i x_i^T) \cdot e^{-y_i x_i^T \omega}\right]^T \\ &= \sum_{i=1}^{m} \frac{e^{-y_i x_i^T \omega}}{1+e^{-y_i x_i^T \omega}} \cdot (-y_i x_i^T)^T \\ &= \sum_{i=1}^{m} \frac{-y_i x_i}{1+e^{y_i x_i^T \omega}} \end{aligned}$$

(d) Notice $\log(1+e^{-y_i x_i^T w}) > 0$, we have

$$\tilde{f}(w) > \frac{\lambda}{2}\|w\|^2 \quad (\lambda > 0)$$

Then

$$\lim_{\|w\| \to \infty} \tilde{f}(w) = +\infty$$

So $\tilde{f}(w)$ is continuous and coercive, it has a global minimum.

This doesn't require the dataset linearly separable, because no assumption about $y_i x_i^T w > 0$ is made.