HOMEWORK ONE

QUESTION 1

(a) .

After transformation, we have:

$$f(x) = (x_1 + x_2)^2 + (x_1 - \frac{1}{2})^2 + 2(x_2 - \frac{1}{2})^2 - \frac{3}{4} \tag{1}$$

We use $||x||_\infty$ as $||x||$ here, then it s.t. $f(x) \to \infty$ as $||x|| \to \infty$. $f(x)$ is thus corercive.

> $\because f(x)$ is also continuous on R^2
>
> $\therefore f(x)$ has a global minimum. (the Corollary of Extreme Value Theorem)

(b).

After transformation, we have:

$$2f(x) = (x_1 + 2x_2)^2 + (x_1 - 1)^2 + 2(x_2 - 1)^2 - 2 \tag{2}$$

We use $||x||_\infty$ as $||x||$ here, then it s.t. $f(x) \to \infty$ as $||x|| \to \infty$. $f(x)$ is thus corercive.

> $\because f(x)$ is also continuous on R^2
>
> $\therefore f(x)$ has a global minimum. (the Corollary of Extreme Value Theorem)

(c).

After transformation, we have:

$$f(x) = (x_1 + x_2 - \frac{1}{2})^2 - x_2 - \frac{1}{4} \tag{3}$$

$$(x_1 + x_2 - \frac{1}{2})^2 - x_2 - \frac{1}{4} \geq -x_2 - \frac{1}{4} \tag{4}$$

Equal could be made when $(x_1 + x_2 - \frac{1}{2}) = 0$

> $\because$ In this case, we then make $x_2 \to \infty$, $(-x_2 - \frac{1}{4}) \to -\infty$
>
> $\therefore f(x)$ doesn' t have a gloabl minmum.

REFERENCE

> Thanks to **Chongxuan Huang** for providing me the method for Q1.(c) . 😁

QUESTION 2

(a).

Let $\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T$

$$f(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{x}||^2 = \frac{1}{2}\sum_{i=1}^{n} x_i^2 \tag{5}$$

$$\begin{aligned} f(\boldsymbol{x})' &= (x_1, x_2, \dots, x_n) \\ \nabla f(\boldsymbol{x}) &= (f(\boldsymbol{x})')^T = \boldsymbol{x} \end{aligned} \tag{6}$$

(b).

$$\begin{aligned} \nabla f(\boldsymbol{w}) &= \frac{1}{2}(X\boldsymbol{w} - \boldsymbol{y})^T(X\boldsymbol{w} - \boldsymbol{y}) + \frac{\lambda}{2}||\boldsymbol{w}||^2 \\ &= \frac{1}{2}(\boldsymbol{w}^T X^T - \boldsymbol{y}^T)(X\boldsymbol{w} - \boldsymbol{y}) + \frac{\lambda}{2}||\boldsymbol{w}||^2 \\ &= \frac{1}{2}(\boldsymbol{w}^T X^T X\boldsymbol{w} - \boldsymbol{w}^T X^T \boldsymbol{y} - \boldsymbol{y}^T X\boldsymbol{w} + \boldsymbol{y}^T\boldsymbol{y}) + \frac{\lambda}{2}||\boldsymbol{w}||^2 \\ &\overset{\boldsymbol{w}^T X^T \boldsymbol{y} = \boldsymbol{y}^T X\boldsymbol{w} \quad (scalar)}{=\!=\!=\!=\!=} \frac{1}{2}(\boldsymbol{w}^T X^T X\boldsymbol{w} - 2\boldsymbol{y}^T X\boldsymbol{w} + \boldsymbol{y}^T\boldsymbol{y}) + \frac{\lambda}{2}||\boldsymbol{w}||^2 \end{aligned} \tag{7}$$

$$\nabla f(\boldsymbol{w}) = X^T X\boldsymbol{w} - X^T\boldsymbol{y} + \lambda\boldsymbol{w} \tag{8}$$

QUESTION 3

(a).

Since there exists a $\boldsymbol{w}_0$ such that

$$y_i \boldsymbol{x}_i^T \boldsymbol{w}_0 > 0, \forall i = 1, 2, 3, \cdots, m. \tag{9}$$

We could make $\boldsymbol{w} = \lambda \boldsymbol{w}_0$,where $\lambda > 0$.

Notice that

$$f(\lambda \boldsymbol{w}_0) = \sum_{i=1}^{m} log(1 + e^{-y_i \boldsymbol{x}_i^T \lambda \boldsymbol{w}_0}) \tag{10}$$

decreases monotonically to zero as $\lambda \to +\infty$, <u>we can reach the conclusion that f doesn't have a global minimum.</u>

(b).

i) Assume $k \in [1, m]$ s.t. $h(\boldsymbol{w}) = -y_k \boldsymbol{x}_k \boldsymbol{w}$

$\because$

$$e^x + 1 > e^x, ln(e^x + 1) > 0 \tag{11}$$

$\therefore$

$$\sum_{i=1}^{m} ln(1 + e^{-y_i \boldsymbol{x}_i \boldsymbol{w}}) \geq ln(1 + e^{-y_k \boldsymbol{x}_k \boldsymbol{w}}) \geq -y_k \boldsymbol{x}_k \boldsymbol{w} = h(\boldsymbol{w}) \tag{12}$$

As a result, we have

$$f(\boldsymbol{w}) \geq h(\boldsymbol{w}) \tag{13}$$

REFERENCE

> Thanks to **Mr. Jiang**, for let me know that "log" here actually means "ln". 😁

---

ii)

> $\because h(\boldsymbol{w})$ is continuous on the compact set S
>
> $\therefore h(\boldsymbol{w})$ has a global minima $\boldsymbol{w}_0$. (Extreme Value Theorem)

And according to the fact that for any $\boldsymbol{w}$, there exists an $i_0 = 1, 2, 3, \ldots, m$ such that

$$y_{i_0}\boldsymbol{x}_{i_0}^T\boldsymbol{w} < 0 \tag{14}$$

We can include that

$$C \triangleq (\boldsymbol{w}_0) = \underset{1\leq i\leq m}{max} - y_i\boldsymbol{x}_i\boldsymbol{w_0} > 0 \tag{15}$$

---

iii)

Since $\forall\boldsymbol{w} \in S$,

$$h(\boldsymbol{w}) \geq h(\boldsymbol{w}_0) = C \tag{16}$$

$||\boldsymbol{w}|| = 1,$ therefore

$$h(\boldsymbol{w}) \geq C||\boldsymbol{w}|| \tag{17}$$

Actually, if $||\boldsymbol{w}|| \neq 1$, we could replace it by $\frac{\boldsymbol{w}}{||\boldsymbol{w}||}$, because

$$\begin{gathered}\underset{1\leq i\leq m}{max} - y_i\boldsymbol{x}_i\boldsymbol{w} \geq C||\boldsymbol{w}|| \\ \Longleftrightarrow \\ \underset{1\leq i\leq m}{max} - y_i\boldsymbol{x}_i\frac{\boldsymbol{w}}{||\boldsymbol{w}||} \geq C\frac{\boldsymbol{w}}{||\boldsymbol{w}||}\end{gathered} \tag{18}$$

Therefore, this holds for $\forall\boldsymbol{w}$.

iv)

Combined with Equation (9), we have

$$f(\boldsymbol{w}) \geq C||\boldsymbol{w}|| \tag{19}$$

Since $f(\boldsymbol{w}) \rightarrow \infty$ as $||\boldsymbol{w}|| \rightarrow \infty$. $f(\boldsymbol{w})$ is thus corercive.

According to the Corollary of Extreme Value Theorem, $f(\boldsymbol{w})$ has a global minimum.

(c).

Let $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^T$

$$\nabla f(\boldsymbol{w}) = ((f(\boldsymbol{w}))')^T = \sum_{i=1}^{m} \frac{-y_i e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}{1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}} \boldsymbol{x_i} \tag{20}$$

(d).

The conclusion is $\tilde{f}(\boldsymbol{w})$ has a global minimum.

We have already proved that $h(\boldsymbol{w})$ has a global minimum in (b).(ii)

And according to (b).(iii) $\forall \boldsymbol{w}$, it holds that

$$h(\boldsymbol{w}) \geq C||\boldsymbol{w}|| \tag{21}$$

Where C could be a negative number, but it won't subvert the conclusion, because we have already added a regularization term to the objective function.

$$\begin{gathered} \tilde{f}(\boldsymbol{w}) \geq C||\boldsymbol{w}|| + \frac{\lambda}{2}||\boldsymbol{w}||_2^2 \\ \lambda > 0 \end{gathered} \tag{22}$$

Since $\tilde{f}(\boldsymbol{w}) \to \infty$ as $||\boldsymbol{w}|| \to \infty$. $\tilde{f}(\boldsymbol{w})$ is thus corercive.

According to the Corollary of Extreme Value Theorem, $\tilde{f}(\boldsymbol{w})$ has a global minimum.

<u>This conclusion doesn't depend on whether the dataset is linearly separable or not.</u>