Problem 1

(a).

It is coercive.
By the *fundamental inequality* we know that $x_1x_2 \geq -(x_1^2+x_2^2)/2$, as is shown in the Hint. Then by substituting x_1x_2 with $-(x_1^2+x_2^2)/2$ we know that $f(\boldsymbol{x}) \geq \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - 3x_1 - 5x_2$, which can be simplified as $f(\boldsymbol{x}) \geq \frac{3}{2}(x_1-1)^2 + \frac{1}{2}(x_2-5)^2 - 14$. In this form it's obvious that as $||\boldsymbol{x}|| \to \infty$, $f(\boldsymbol{x}) \to +\infty$. Thus $f(\boldsymbol{x})$ is coercive.

(b).

Minimum is $-44/7$.
Maximum does not exist.
Since $f(\boldsymbol{x})$ is coercive, the maximum does not exist. By the continuity of f we know that minimum must exist. Then we take the $\nabla f(\boldsymbol{x})^T$, in this case $(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})$. By letting $\nabla f(\boldsymbol{x})^T = \boldsymbol{0}$ and solving the equation we get $(x_1, x_2) = (1/7, 17/7)$. We still need to calculate the *Hessian Matrix* which in this case is $H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$, then by solving the second derivative we can see that $H = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$, then $det(H) = 7 > 0$, which tells us that $(1/7, 17/7)$ actually is the minimum point of $f(\boldsymbol{x})$. By substituting (x_1, x_2) with $(1/7, 17/7)$ we get the minimum value $-44/7$.

Problem 2

(a).

It does not have a global minimum.
We can prove this by contradiction. Assume it does have a global minimum, denoted by $f(\boldsymbol{w_*})$, we will discuss it in two cases.

1. $\boldsymbol{w_*}$ itself satisfies that $y_i\boldsymbol{x_i^T w_*} > 0$, $\forall i = 1, 2, \ldots, m$.
   In this case simply take $2\boldsymbol{w_*}$, it's easy to see that $y_i\boldsymbol{x_i^T 2 w_*} = 2y_i\boldsymbol{x_i^T w_*} > y_i\boldsymbol{x_i^T w_*} > 0$, $\forall i = 1, 2, \ldots, m$. Since $log(1+e^{-z_i})$ where $z_i = y_i\boldsymbol{x_i^T w}$ is monotonically decreasing with z_i, $f(2\boldsymbol{w_*})$ must be smaller than $f(\boldsymbol{w_*})$. This contradicts our assumption that $f(\boldsymbol{w_*})$ is the global minimum.

2. $\exists i$ such that $y_i \boldsymbol{x_i^T} \boldsymbol{w_*} \leq 0$.
   Now we must use $f(\boldsymbol{w_0})$. Since $y_i \boldsymbol{x_i^T} \boldsymbol{w_0} > 0$, $\forall i = 1, 2, \ldots, m$, for those i such that $y_i \boldsymbol{x_i^T} \boldsymbol{w_*} > 0$, it's easy to find an $\alpha > 0$ which makes $\alpha y_i \boldsymbol{x_i^T} \boldsymbol{w_0} > y_i \boldsymbol{x_i^T} \boldsymbol{w_*} > 0$. And note that this α also satisfies that $\alpha y_i \boldsymbol{x_i^T} \boldsymbol{w_0} > 0 \geq y_i \boldsymbol{x_i^T} \boldsymbol{w_*}$ for those i such that $y_i \boldsymbol{x_i^T} \boldsymbol{w_*} \leq 0$. This means that $\alpha y_i \boldsymbol{x_i^T} \boldsymbol{w_0} > y_i \boldsymbol{x_i^T} \boldsymbol{w_*}$, $\forall i = 1, 2, \ldots, m$. Then by the monotonicity again $f(\alpha \boldsymbol{w_0}) < f(\boldsymbol{w_*})$. This also contradicts our assumption.

To summarize, f does not have a global minimum.

(b).

i)

Since there exists an $i_0 = 1, 2, \ldots, m$ such that $y_{i_0} \boldsymbol{x_{i_0}^T} \boldsymbol{w} < 0$ for any $\boldsymbol{w}$ and $h(\boldsymbol{w})$ is the maximum of $-y_i \boldsymbol{x_i^T} \boldsymbol{w}$ for $1 \leq i \leq m$, the definition of $h(\boldsymbol{w})$ itself satisfies that $h(\boldsymbol{w}) > 0$ for any $\boldsymbol{w}$. And it's obvious that $log(1 + e^z) > log(e^z) = z$ for $z > 0$. For any $\boldsymbol{w}$, let z_i denote $\max\limits_{1 \leq i \leq m} -y_i \boldsymbol{x_i^T} \boldsymbol{w}$. Then $h(\boldsymbol{w}) = z_i > 0$. Then $f(\boldsymbol{w}) > log(1 + e^{z_i}) > z_i = h(\boldsymbol{w})$.

ii)

It's obvious that S is both bounded and closed, thus in our problem it is compact. And this problem allows us to assume that h is continuous. Then by Extreme Value Theorem $h(\boldsymbol{w})$ has a global minimum $\boldsymbol{w_0}$ on S.

iii)

For any $\boldsymbol{w}$, its obvious that $\frac{\boldsymbol{w}}{||\boldsymbol{w}||} \in S$, where S is the set defined in ***ii*).** Then by definition $h(\frac{\boldsymbol{w}}{||\boldsymbol{w}||}) \geq C$. Yet $h(\frac{\boldsymbol{w}}{||\boldsymbol{w}||})$ is simply $\frac{h(\boldsymbol{w})}{||\boldsymbol{w}||}$, thus $h(\boldsymbol{w}) \geq C||\boldsymbol{w}||$ for any $\boldsymbol{w}$.

iv)

From ***i*).** and ***iii*).** we know $f(\boldsymbol{w}) \geq C||\boldsymbol{w}||$ for any $\boldsymbol{w}$, thus $f(\boldsymbol{w}) \to +\infty$ as $\boldsymbol{w} \to \infty$. Thus f is coercive, which means it has a global minimum.

(c).

This question requires us to calculate the gradient of $f(\boldsymbol{w})$, we can calculate it directly. We write

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

as

$$\sum_{i=1}^{m} log(1+e^{-y_i \sum_{j=1}^{n} x_{ij} w_j})$$

then we take the partial derivative of $w_k, k = 1, 2, \dots, n$.

$$\frac{\partial f}{\partial w_k} = \sum_{i=1}^{m} \frac{-y_i x_{ik} e^{-y_i \boldsymbol{x_i^T w}}}{1+e^{-y_i \boldsymbol{x_i^T w}}}$$

Finally we get

$$\frac{\partial f}{\partial \boldsymbol{w}} = \sum_{i=1}^{m} \frac{-y_i \boldsymbol{x_i} e^{-y_i \boldsymbol{x_i^T w}}}{1+e^{-y_i \boldsymbol{x_i^T w}}}$$

or equivalently

$$\nabla f(\boldsymbol{w}) = (\sum_{i=1}^{m} \frac{-y_i \boldsymbol{x_i} e^{-y_i \boldsymbol{x_i^T w}}}{1+e^{-y_i \boldsymbol{x_i^T w}}})^T$$

Problem 3

(a).

The first-order Taylor Expansion with Lagrange remainder for univariate function $g(s)$ at point a is $g(a+s) = g(a)+g'(a)s+\frac{1}{2}g''(a+ts)s^2$ for some $t \in (0,1)$. If we apply this function at point 0 we get $g(s) = g(0)+g'(0)s+\frac{1}{2}g''(ts)s^2$. Let $f(\boldsymbol{x}+s\boldsymbol{\hat{d}}) = g(s)$ where $\boldsymbol{\hat{d}} = \boldsymbol{d}/||\boldsymbol{d}||$. By applying the univariate Taylor Expansion to it we have $f(\boldsymbol{x}+s\boldsymbol{\hat{d}}) = g(0)+g'(0)s+\frac{1}{2}g''(ts)s^2$. $g(0)$ is simply $f(\boldsymbol{x})$. And by applying the chain rule we derive that $g'(0) = \nabla f(\boldsymbol{x})^T\boldsymbol{\hat{d}}$, $g''(ts) = \boldsymbol{\hat{d}^T}\nabla^2 f(\boldsymbol{x}+ts\boldsymbol{\hat{d}})\boldsymbol{\hat{d}}$. Thus $f(\boldsymbol{x}+s\boldsymbol{\hat{d}}) = f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T\boldsymbol{\hat{d}}s+\frac{1}{2}\boldsymbol{\hat{d}^T}\nabla^2 f(\boldsymbol{x}+ts\boldsymbol{\hat{d}})\boldsymbol{\hat{d}}s^2$. Let $s = ||\boldsymbol{d}||$ we get $f(\boldsymbol{x}+\boldsymbol{d}) = f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T\boldsymbol{d}+\frac{1}{2}\boldsymbol{d^T}\nabla^2 f(\boldsymbol{x}+t\boldsymbol{d})\boldsymbol{d}$ where $t \in (0,1)$.

(b).

If we take the derivative of $\nabla f(\boldsymbol{x}+t\boldsymbol{d})$ treating t as the variable, by Chain rule we get $\nabla^2 f(\boldsymbol{x}+t\boldsymbol{d})\boldsymbol{d}$. Thus the integral $\int_0^1 \nabla^2 f(\boldsymbol{x}+t\boldsymbol{d})\boldsymbol{d}\,dt$ is simply $\nabla f(\boldsymbol{x}+t\boldsymbol{d})|_0^1 = \nabla f(\boldsymbol{x}+\boldsymbol{d}) - \nabla f(\boldsymbol{x})$, thus $\nabla f(\boldsymbol{x}+\boldsymbol{d}) = \nabla f(\boldsymbol{x}) + \int_0^1 \nabla^2 f(\boldsymbol{x}+t\boldsymbol{d})\boldsymbol{d}\,dt$.

Problem 4

We solve this problem by calculating the eigenvalues of the matrices.
The eigenvalues of A are $\lambda_1 = 8, \lambda_2 = 2, \lambda_3 = 5$. Thus A is positive definite.
The eigenvalues of B are $\lambda_1 = 3.656, \lambda_2 = -3.233, \lambda_3 = -0.423$. Thus B is indefinite.
The eigenvalues of C are $\lambda_1 = 3, \lambda_2 = 3, \lambda_3 = 0$. Thus C is positive semidefinite.