Problem 1:

a)

$$\begin{aligned}\sigma'(x) &= -(1+e^{-x})^{-2}(-e^{-x})\\ &= \frac{1}{1+e^{-x}}\frac{e^{-x}}{1-e^{-x}}\\ &= \frac{1}{1+e^{-x}}\left(1-\frac{1}{1-e^{-x}}\right)\\ &= \sigma(x)(1-\sigma(x))\end{aligned}$$

Since $0 < \sigma(x) < 1$ for all x, $0 < 1-\sigma(x) < 1$ as well, and $\sigma'(x) > 0$. b)

$$\begin{aligned}\frac{\partial \ell_i}{\partial a} &= -y_i\frac{\sigma'(a+bx_i)}{\sigma(a+bx_i)} + (1-y_i)\frac{\sigma'(a+bx_i)}{1-\sigma(a+bx_i)}\\ &= -y_i(1-\sigma(a+bx_i)) + (1-y_i)\sigma(a+bx_i)\\ &= \sigma(a+bx_i) - y_i,\\ \frac{\partial \ell_i}{\partial b} &= -y_i\frac{x_i\sigma'(a+bx_i)}{\sigma(a+bx_i)} + (1-y_i)\frac{x_i\sigma'(a+bx_i)}{1-\sigma(a+bx_i)}\\ &= -y_ix_i(1-\sigma(a+bx_i)) + (1-y_i)x_i\sigma(a+bx_i)\\ &= x_i(\sigma(a+bx_i) - y_i),\end{aligned}$$

Therefore,

$$\nabla \ell_i(a,b) = (\sigma(a+bx_i) - y_i)\begin{bmatrix}1\\ x_i\end{bmatrix}.$$

$$\begin{aligned}\frac{\partial^2 \ell_i}{\partial a^2} &= \sigma'(a+bx_i),\\ \frac{\partial^2 \ell_i}{\partial b^2} &= x_i^2\sigma'(a+bx_i), \frac{\partial^2 \ell_i}{\partial a \partial b} &= x_i\sigma'(a+bx_i),\end{aligned}$$

Thus,

$$\nabla^2 \ell_i(a,b) = \sigma'(a+bx_i) = \begin{bmatrix}1 & x_i\\ x_i & x_i^2.\end{bmatrix}$$

The eigenvalues of this matrix are 0 and $1+x_i^2$, both nonnegative. Thus, $\nabla^2 \ell_i(a,b)$ is positive semidefinite.

c)

The nullspace of $\nabla^2 \ell_i(a,b)$ has dimension 1, equal to the multiplicity of 0 as an eigenvalue. Therefore, the nonzero vector $\begin{bmatrix}x_i & -1\end{bmatrix}^T$ is a basis of $N(\nabla^2 \ell_i(a,b))$.

d)

Since $x_i \neq x_j$, the vectors $\begin{bmatrix} x_i & -1 \end{bmatrix}^T$ and $\begin{bmatrix} x_j & -1 \end{bmatrix}^T$ are linearly independent. Therefore, the subspaces they span, which, by (c), are the nullspaces of $\nabla^2 \ell_i(a,b)$ and $\nabla^2 \ell_j(a,b)$, respectively, intersect trivially. Therefore,

$$\{\mathbf{0}\} \subseteq \bigcap_{k=1}^{n} N(\nabla^2 \ell_k(a,b)) \subseteq N(\nabla^2 \ell_i(a,b)) \cap N(\nabla^2 \ell_j(a,b)) = \{\mathbf{0}\}.$$

e)

Suppose $\boldsymbol{x}^T \nabla^2 \ell(a,b) \boldsymbol{x} = 0$ for some $\boldsymbol{x} \in \mathbb{R}^2$. We must show that $\boldsymbol{x} = 0$.

Since $\ell = \sum \ell_i$,

$$\boldsymbol{x}^T \nabla^2 \ell(a,b) \boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{x}^T \nabla^2 \ell_i(a,b) \boldsymbol{x}.$$

As $\nabla^2 \ell_i(a,b)$ is positive semidefinite for all i, by (b), $\boldsymbol{x}^T \nabla^2 \ell_i(a,b) \boldsymbol{x} \geq 0$ for all i. It follows that

$$\boldsymbol{x}^T \nabla^2 \ell_i(a,b) \boldsymbol{x} = 0$$

for all i.

Fact: If A is a positive-semidefinite, symmetric matrix, there is a matrix B of real numbers such that $A = B^T B$. Moreover, $N(A) = N(B)$ (and $C(A) = C(B^T)$). (This result follows from the Spectral Theorem.)

Write

$$\nabla^2 \ell_i(a,b) = B_i^T B_i.$$

Then

$$0 = \boldsymbol{x}^T \nabla^2 \ell_i(a,b) \boldsymbol{x} = \boldsymbol{x}^T B_i^T B_i \boldsymbol{x} = (B_i \boldsymbol{x})^T (B_i \boldsymbol{x}) = \|B_i \boldsymbol{x}\|^2,$$

and it follows that $B_i \boldsymbol{x}_i = \mathbf{0}$ for all i. So

$$\boldsymbol{x} \in N(B_i) = N(\nabla^2 \ell_i(a,b))$$

and, thus,

$$\boldsymbol{x} \in \bigcap_{i=1}^{n} N(\nabla^2 \ell_i(a,b)).$$

This intersection is zero by our hypothesis that the x_i are not all equal, and by (d). Therefore, $\boldsymbol{x} = \mathbf{0}$ and $\nabla^2 \ell(a,b)$ is positive definite.

(f)

Since $\nabla^2 \ell(a,b)$ is positive definite, $\ell(a,b)$ is strictly convex. Therefore, $\ell(a,b)$ has at most one local minimum.

Problem 2.

a)

Since $0 < \sigma(x) < 1$ for all x, it followews that $0 < p_i(a,b) < 1$ for all a, b. Therefore, $\log p_i(a,b) < 0$ and $\ell_i(a,b) = -\log p_i(a,b) > 0$.

b)

Suppose

$$\lim_{t\to\infty} \ell_i(tv_1, tv_2) = \infty.$$

Then

$$\lim_{t\to\infty} \log p_i(tv_1, tv_1) = -\lim_{t\to\infty} \ell_i(tv_1, tv_2) = -\infty.$$

It follows that

$$\lim_{t\to\infty} p_i(tv_1, tv_2) = 0.$$

Suppose, first, that $y_i = 1$. Then

$$p_i(tv_1, tv_2) = \sigma(t(v_1 + v_2 x_i))$$

and

$$\lim_{t\to\infty} \sigma(t(v_1 + v_2 x_i)) = 0.$$

This happens when $t(v_1 + v_2 x_i) \to -\infty$ as $t \to \infty$, i.e., when

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ x_1 \end{bmatrix} = v_1 + v_2 x_i < 0,$$

i.e., when

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in H\left(-\begin{bmatrix} 1 \\ x_i \end{bmatrix} \right).$$

Suppose, now, that $y_i = 0$, so that $p_i(tv_1, tv_2) = 1 - \sigma(t(v_1 + v_2 x_i))$. We have

$$\lim_{t\to\infty} (1 - \sigma(t(v_1 + v_2 x_i))) = 0$$

if and only if

$$\lim_{t\to\infty} \sigma(t(v_1 + v_2 x_i)) = 1$$

if and only if $t(v_1 + v_2 x_i) \to +\infty$ as $t \to \infty$, i.e., when

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ x_1 \end{bmatrix} = v_1 + v_2 x_i > 0,$$

i.e., when

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in H\left(\begin{bmatrix} 1 \\ x_i \end{bmatrix} \right).$$

Setting

$$H_i = \begin{cases} H(\boldsymbol{w}_i) & \text{if } y_i = 0, \\ H(-\boldsymbol{w}_i) & \text{if } y_i = 1 \end{cases}, \quad \text{where} \quad \boldsymbol{w}_i = \begin{bmatrix} 1 \\ x_i \end{bmatrix}.$$

d)

Suppose $y_i = y_k = 0$ and $y_j = 1$. Then

$$H_i = H(\boldsymbol{w}_i), \quad H_j = H(-\boldsymbol{w}_j), \quad \text{and} \quad H_k = H(\boldsymbol{w}_k)$$

Note that

$$-(-\boldsymbol{w}_j) = \boldsymbol{w}_j = a\boldsymbol{w}_i + b\boldsymbol{w}_k, \quad \text{where} \quad a = \frac{x_k - x_j}{x_k - x_i} \quad \text{and} \quad a = \frac{x_j - x_i}{x_k - x_i}.$$

Using the inequalities $x_1 < x_2 < x_3$, one deduces that $0 < a, b < 1$. Therefore, by (c),

$$H_i \cup H_j \cup H_k = \mathbb{R}^2 - \{\mathbf{0}\}.$$

The case where $y_i = y_k = 1$ and $y_j = 0$ is argued similarly.

e)

Let $\boldsymbol{v} \in \mathbb{R}^2$ be nonzero. Then $\boldsymbol{v}$ belongs to one of H_i, H_j, or H_k. Without loss of generality, suppose $\boldsymbol{v} \in H_i$. Then, by (b),

$$\lim_{t\to\infty} \ell_i(tv_1, tv_2) = \infty.$$

Since each $\ell(a, b) \geq \ell_i(a, b)$ by (a), we have

$$\lim_{t\to\infty} \ell_i(tv_1, tv_2) = \infty,$$

too. Therefore, that S_K does not contain the ray $\{t\boldsymbol{v} : t \geq 0\}$. Since $\boldsymbol{v}$ was arbitrary, S_K doesn't contain any ray of this form.

f)

By Problem 1(b), $\ell(a, b)$ is convex. Therefore S_K is convex. Since S_K is convex and, by e), contains no ray, S_K is bounded. Being closed and bounded, it's compact.

g)

Since ℓ is continuous and S_K is compadt, ℓ achieves achieves its infimum on S_K at $(\widehat{a}, \widehat{b}) \in S_K$, say. If $\ell(a, b) \leq \ell(\widehat{a}, \widehat{b})$, then $\ell(a, b) \leq K$ and $(a, b) \in K$. It follows that $\ell(a, b) = \ell(\widehat{a}, \widehat{b})$. Therefore,

$$\ell(\widehat{a}, \widehat{b}) = \inf_{(a,b)\in\mathbb{R}^2} \ell(a, b).$$

We've assumed our x_i values are all distinct, so the hypothesis of 1(f) holds. Therefore, $(\widehat{a}, \widehat{b})$ is the unique point at which ℓ achieves its minimum value.

Let $A_i = \nabla^2 \ell_i(a, b)$ and let $A = \nabla^2 \ell(a, b) = \sum A_i$.

(1b) Show that A_i has two nonnegative eigenvalues.

(1c) Find a basic solution of the system

$$A_i \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

(1d) An element of $N(A_i) \cap N(A_j)$ is a solution of the homogeneous system

$$\begin{bmatrix} A_i \\ A_j \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Use the fact that $x_i \neq x_j$ to deduce that this system has only the trivial solution.

(1e) Let $\boldsymbol{x}$ be such that $\boldsymbol{x}^T A \boldsymbol{x} = 0$. We need to show that $\boldsymbol{x} = 0$. Note that

$$\boldsymbol{x}^T A \boldsymbol{x} = \sum \boldsymbol{x}^T A_i \boldsymbol{x}.$$

The $\boldsymbol{x}^T A_i \boldsymbol{x}$ are nonnegative (why?), so $\boldsymbol{x}^T A \boldsymbol{x} = 0$ if and only if $\boldsymbol{x}^T A_i \boldsymbol{x} = 0$ for all i. By the Lemma below, this holds if and only if $\boldsymbol{x} \in N(A_i)$ for all i...

Lemma: If S is a positive semidefinite, symmetric matrix, then

$$\{\boldsymbol{y} : \boldsymbol{y}^T S \boldsymbol{y} = 0\} = N(S)$$

Proof: Clearly, $N(S) \subseteq \{\boldsymbol{y} : \boldsymbol{y}^T S \boldsymbol{y} = 0\}$. Conversely, suppose $\boldsymbol{y}^T S \boldsymbol{y} = 0$. Since S is symmetric and positive semidefinite, there is a matrix R of real numbers such that $S = R^T R$. (This follows from the the Spectral Theorem.) Then

$$0 = \boldsymbol{y}^T S \boldsymbol{y} = \boldsymbol{y}^T R^T R \boldsymbol{y} = (R\boldsymbol{y})^T (R\boldsymbol{y}) = \|R\boldsymbol{y}\|^2,$$

so $R\boldsymbol{y}$ must be zero. Thus, $\boldsymbol{y} \in N(R) \subseteq N(S)$.

(2a) Use the fact that $0 < \sigma(x) < 1$ for all x.

(2b) Show that

$$\lim_{t \to \infty} \ell_i(tv_1, tv_2) = \infty$$

if and only if

- $(v_1 + v_2 x_1) > 0$, if $y_i = 0$.
- $(v_1 + v_2 x_1) < 0$, if $y_i = 1$.

(2c) I'll accept some nicely drawn pictures illustrating what's going on here in lieu of a formal proof.

(2d) Suppose $y_i = y_k = 0$ and $y_j = 1$. By (b),

$$H_i = H(\boldsymbol{w}_i), \quad H_j = H(-\boldsymbol{w}_j), \quad \text{and} \quad H_k = H(\boldsymbol{w}_k),$$

where $\boldsymbol{w}_i = \begin{bmatrix} 1 & x_i \end{bmatrix}^T$. Show that $\boldsymbol{w}_j = a\boldsymbol{w}_i + b\boldsymbol{w}_k$ with $a, b > 0$. Then invoke (c).

2(e) Use (b).