1.

a) $f: \mathbb{R}^n \mapsto \mathbb{R}$, $f(x) = \|x\|$

$$f(\vec{x}) = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

$$\nabla f(\vec{x}) = \begin{bmatrix} x_i \left(\sum_{i=1}^{n} x_i^2 \right)^{-\frac{1}{2}} \\ \\ x_n \left(\sum_{i=1}^{n} x_i^2 \right)^{-\frac{1}{2}} \end{bmatrix}$$

$$\|\nabla f(\vec{x})\| = \left(\left(\sum_{i=1}^{n} x_i^2 \right)^{-1} \sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

$$= \left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2} \right)^{1/2} = 1 \leq 1$$

Answer : $L = 1$

b) $f: \mathbb{R}^n \mapsto \mathbb{R}$, $f(x) = \|\vec{x}\|^2$

$$f(\vec{x}) = \sum_{i=1}^{n} x_i^2$$

$$\nabla f(\vec{x}) = \begin{bmatrix} 2x_1 & 2x_2 & \cdots & 2x_n \end{bmatrix}^T$$

$$\|\nabla f(\vec{x})\| = \left(4\sum_{i=1}^{n} x_i^2\right)^{1/2} = 2\left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

$\|\nabla f(\vec{x})\| = 2 \cdot \|\vec{x}\|$ when any x_i in the vector goes to ∞ the gradient will also go to ∞. which means we can't find a L such that

$$\|\nabla f(\vec{x})\| \leq L$$

c) $\rho : \mathbb{R} \mapsto \mathbb{R}$, $\rho(x) = \dfrac{1}{1+e^{-x}} \leq 1$

$$\rho'(x) = \rho(x)(1-\rho(x)) = \rho(x) - \rho(x)^2$$

$$\rho'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})(1+e^{-x})} =$$

We want x to be big since the denominator grows faster than the numerator.

$$\lim_{x \to -\infty} \frac{e^{-x}}{(1+e^{-x})^2} \leq \lim_{x \to \infty} \frac{e^{-x}}{(1+e^{-x})^2} = 1 = L$$

$$\frac{e^{-x}}{(1+e^{-x})^2} \leq L = 1$$

1 d) $f: \mathbb{R}^n \mapsto \mathbb{R}$ for $f(x) = \rho(w^T x + b)$

where $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ and ρ is from (c).

$$\nabla f(x) = \left[w_1 \frac{e^{-(w^T x + b)}}{(1+e^{-(w^T x+b)})^2} \cdots w_n \frac{e^{-(w^T x+b)}}{(1+e^{-(w^T x+b)})^2} \right]^T$$

$$\|\nabla f(x)\| = \left(\frac{e^{-2(w^T x+b)}}{(1+e^{-(w^T x+b)})^4} \cdot \sum_{i=0}^{n} w_i^2 \right)^{1/2}$$

$$\|\nabla f(x)\| = \frac{e^{-(w^T x+b)}}{(1+e^{-(w^T x+b)})^2} \cdot \|w\|$$

$$= \left(\frac{1 + e^{-(w^T x+b)}}{(1+e^{-(w^T x+b)})^2} - \frac{1}{(1+e^{-(w^T x+b)})^2} \right) \|w\|$$

$$= \left(\frac{1}{1+e^{-(w^T x+b)}} - \frac{1}{(1+e^{-(w^T x+b)})^2} \right) \cdot \|w\|$$

1 d)

we know that:

$$(*) \quad \frac{1}{1+e^{-(w^T x+b)}} > \frac{1}{(1+e^{-(w^T x+b)})^2}$$

- Since $\boxed{1+e^{-(w^T x+b)} > 1}$, so when we raise $1+e^{-(w^T x+b)}$ to the power of 2:

$$1+e^{-(w^T x+b)} < (1+e^{-(w^T x+b)})^2 \Rightarrow (*)$$

- we also know both sides of $(*)$ are greater than zero.

- we therefore know:

$$\|\nabla f(x)\| = \underbrace{\left(\frac{1}{1+e^{-(w^T x+b)}} - \frac{1}{(1+e^{-(w^T x+b)})^2} \right)}_{<1} \cdot \|w\| \leq \|w\|$$

<u>Answer</u>: $L = \|w\|$

2a)

$$\left\| f\left(\frac{1}{T}\sum_{s=0}^{s-1} x_s\right) - f(x^*) \right\| \leq \frac{L \cdot R}{\sqrt{T}} = 10^{-4}$$

$$T = 15^2 \cdot 10^8 \rightarrow \boxed{T = 2{,}25 \cdot 10^{10}}$$

$$\eta = \frac{R}{L \cdot \sqrt{T}} \rightarrow \eta = \frac{5}{3 \cdot 15 \cdot 10^4} \rightarrow \boxed{\eta = \frac{1}{9} \cdot 10^{-4}}$$

Answer: $\boxed{\eta = \frac{1}{9} \cdot 10^{-4}}$ and

$(2{,}25 \cdot 10^{10} - 1)$ number of iterations.

2
b)

$$f(x^{(t)}) - f(x^*) \leq \frac{\|x_0 - x^*\|^2}{2\eta \cdot t} \leq 10^{-4}$$

let $\eta = \frac{1}{3}$

$$\frac{25}{2 \cdot \frac{1}{3} \cdot 10^{-4}} \leq t \rightarrow 3{,}75 \cdot 10^5 \leq t$$

Answer: $\eta = \frac{1}{3}$ and $t \geq 3{,}75 \cdot 10^5$

3a)

<u>Theorem</u>: where $f: \Omega_0 \mapsto \mathbb{R}$ and finding $\boxed{\min_{x \in \Omega} f(x)}$ is a convex optimization problem (f & Ω are both convex) and $\Omega \subseteq \Omega_0$ is a convex set.

Then if x^* is a local minimizer, x^* is also a global minimizer.

I assume the domain of the function is $\mathbb{R}^2$ since nothing else is given. This means the domain is convex.

If $f: \mathbb{R}^n \mapsto \mathbb{R}$ is twice continuously differentiable, then f is convex iff:

$$0 \preceq \nabla^2 f(x) \quad \forall x \in \mathbb{R}^n$$

3a) $f(x_1, x_2) = (2x_1 - 1)^4 + (x_1 + x_2 - 1)^2$

$$\nabla f(\vec{x}) = \begin{bmatrix} 8(2x_1-1)^3 + 2(x_1+x_2-1) \\ 2(x_1+x_2-1) \end{bmatrix}$$

$$\nabla^2 f(\vec{x}) = \begin{bmatrix} 48(2x_1-1)^2 + 2 & 2 \\ 2 & 2 \end{bmatrix}$$

- Let v be some vector s.t. $v \in \mathbb{R}^n$

$$v^T \nabla^2 f(\vec{x}) v = v^T \begin{bmatrix} (48(2x_1-1)^2+2)v_1 + 2v_2 \\ 2v_1 + 2v_2 \end{bmatrix}$$

$$v^T \nabla^2 f(\vec{x}) \cdot v = 48v_1^2(2x_1-1)^2 + 2v_1^2 + 4v_1v_2 + 2v_2^2$$

$$= 48v_1^2(2x_1-1)^2 + (\sqrt{2}v_1 + \sqrt{2}v_2)^2 \geq 0$$

$\Rightarrow$ $\boxed{\text{Hessian of } f \text{ is PSD for all possible } \vec{x}, \vec{v} \in \mathbb{R}^2}$ $\Rightarrow$ f is convex

3 a)

$$\nabla f(x) = 0 \rightarrow \begin{cases} 8(2x_1-1)^3 + 2(x_1+x_2-1) = 0 \\ 2(x_1+x_2-1) = 0 \end{cases}$$

$$\begin{cases} 2x_1 - 1 = 0 \quad (1) \\ x_1 + x_2 - 1 = 0 \quad (2) \end{cases}$$

$(1) \Rightarrow \boxed{x_1 = \frac{1}{2}}$

$(2) \& \left[x_1 = \frac{1}{2}\right] \Rightarrow \boxed{x_2 = -\frac{1}{2}}$

Answer: $x_1 = \frac{1}{2}$, $x_2 = -\frac{1}{2}$

5 b)

(i)

$$\begin{cases} x_1^{(1)} = \mu^{(0)} \cdot 10 \\ x_2^{(1)} = \mu^{(0)} \cdot 2 \end{cases}$$

-10

-2

(ii) let $\boxed{\mu^{(0)} = 1}$

① $f(10,2) \geq f(0,0) - 0{,}2 \cdot \|\nabla f(0,0)\|^2$

② let $\mu^{(0)} \leftarrow \mu^{(0)} \cdot \gamma = 0{,}2$

$f(2;0.4) > f(0,0) \cdot 0{,}2 \cdot 0{,}2 \cdot \|\nabla f(0,0)\|^2$

③ let $\mu^{(0)} \leftarrow \mu^{(1)} \cdot 0{,}2 = 0{,}04$

$f(0.4;0.08) \leq f(0,0) \cdot 0{,}04 \cdot 0{,}2 \cdot \|\nabla f(0,0)\|^2$

The Amijo condition holds after iteration three.

Answer: $\mu^{(0)} = 0{,}04$ satisfies the Amijo condition

3 b)

(iii)

We want $\mu^{(0)} \leq 0{,}04$

$$\mu_0 \cdot \beta^n \leq 0{,}04 \rightarrow 1 \cdot \left(\frac{1}{2}\right)^n \leq 0{,}04$$

$$-n \cdot \log(2) \leq \log\left(\frac{4}{100}\right)$$

$$-n \leq \frac{\log\left(\frac{4}{100}\right)}{\log(2)}$$

$$n \geq -\frac{\log\left(\frac{4}{100}\right)}{\log(2)} = 2\log_2(5) \approx 4{,}6$$

Answer: We might have to backtrack 5 times.