Question 1.1

Let $w^* = (1,0)$, $w = (0,1)$, $\epsilon = \sqrt{2}$, $x = (1,-1)$, $S = \{(x,1)\}$. For any w' s.t. $\|w - w'\| \leq \epsilon$, it is clear that we have $L_S[w] = L_S[w'] = 1$. Hence, it holds that $L_S[w] \leq L_S[w']$. Moreover, $L_S[w^*] = 0$ as it classifies x correctly:

$$\begin{aligned} f_{w^*}(x) = \langle x, w^* \rangle = (1 \cdot 1) + (-1 \cdot 0) = 1 &\Rightarrow y f_{w^*}(x) = 1 \\ &\Rightarrow l(f_{w^*}(x), y) = 0 \end{aligned}$$

Therefore, w is a local minima but not a global minima, as required. $\square$

SGD proof of lemma 1

$$\begin{aligned}
\sum_{t=1}^{T} \left\langle w^{(t)} - w^*, v_t \right\rangle &= \sum_{t=1}^{T} \frac{1}{\mu} \left\langle w^{(t)} - w^*, \mu v_t \right\rangle \\
&= \sum_{t=1}^{T} \frac{1}{2\mu} \left(- \left\| w^{(t)} - w^* - \mu v_t \right\|^2 + \left\| w^{(t)} - w^* \right\|^2 + \mu^2 \left\| v_t \right\|^2 \right) \\
&= \sum_{t=1}^{T} \frac{1}{2\mu} \left(- \left\| w^{(t)} - w^* - \left(w^{(t)} - w^{(t+1)} \right) \right\|^2 + \left\| w^{(t)} - w^* \right\|^2 + \mu^2 \left\| v_t \right\|^2 \right) \\
&= \frac{1}{2\mu} \sum_{t=1}^{T} \left(- \left\| w^{(t+1)} - w^* \right\|^2 + \left\| w^{(t)} - w^* \right\|^2 \right) + \frac{\mu}{2} \sum_{t=1}^{T} \left\| v_t \right\|^2 \\
&= \frac{1}{2\mu} \left(- \left\| w^{(t+1)} - w^* \right\|^2 + \left\| w^{(1)} - w^* \right\|^2 \right) + \frac{\mu}{2} \sum_{t=1}^{T} \left\| v_t \right\|^2 \\
&= \frac{1}{2\mu} \left(- \left\| w^{(t+1)} - w^* \right\|^2 + \left\| 0 - w^* \right\|^2 \right) + \frac{\mu}{2} \sum_{t=1}^{T} \left\| v_t \right\|^2 \\
&\leq \frac{1}{2\mu} \left\| w^* \right\|^2 + \frac{\mu}{2} \sum_{t=1}^{T} \left\| v_t \right\|^2
\end{aligned}$$

SGD proof of lemma 2 (using lemma 1)

$$
\begin{aligned}
\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{T}\sum_{t=1}^{T}\left\langle w^{(t)}-w^*,v_t\right\rangle\right] &= \frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\sum_{t=1}^{T}\left\langle w^{(t)}-w^*,v_t\right\rangle\right]\\
&\leq \frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{2\mu}\|w^*\|^2+\frac{\mu}{2}\sum_{t=1}^{T}\|v_t\|^2\right]\\
&= \frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{2\mu}\|w^*\|^2\right]+\frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{\mu}{2}\sum_{t=1}^{T}\|v_t\|^2\right]\\
&\leq \frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{2\mu}B^2\right]+\frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{\mu}{2}\sum_{t=1}^{T}\rho^2\right]\\
&= \frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{\rho\sqrt{T}}{2B}B^2\right]+\frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{B}{2\rho\sqrt{T}}T\rho^2\right]\\
&= \frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{\rho\sqrt{T}}{2}B\right]+\frac{1}{T}\mathbb{E}_{v_1,\ldots,v_T}\left[\frac{B\sqrt{T}\rho}{2}\right]\\
&= \frac{1}{T}\frac{\rho\sqrt{T}}{2}B+\frac{1}{T}\frac{\rho\sqrt{T}}{2}B\\
&= \frac{B\rho}{\sqrt{T}}
\end{aligned}
$$

SGD proof of lemma 3

Due to the convexity of g, it holds that

$$g(w^{(t)})-g(w^*)\leq\left\langle w^{(t)}-w^*,\nabla g(w^{(t)})\right\rangle=\left\langle w^{(t)}-w^*,v_t\right\rangle$$

Hence

$$\sum_{t=1}^{T}\mathbb{E}_{v_t}\left[g(w^{(t)})-g(w^*)\right]\leq\sum_{t=1}^{T}\mathbb{E}_{v_t}\left[\left\langle w^{(t)}-w^*,\nabla g(w^{(t)})\right\rangle\right]$$

Therefore, using the linearity of expected value:

$$\mathbb{E}_{v_1,\ldots,v_T}\left[\sum_{t=1}^{T}\left(g(w^{(t)})-g(w^*)\right)\right]\leq\mathbb{E}_{v_1,\ldots,v_T}\left[\sum_{t=1}^{T}\left\langle w^{(t)}-w^*,\nabla g(w^{(t)})\right\rangle\right]$$

Let's conclude

By Jensen's Inequality:

$$\begin{aligned}
\mathbb{E}_{v_1,\ldots,v_T}\left[g(\bar{w})\right] - g(w^*) &= \mathbb{E}_{v_1,\ldots,v_T}\left[g\left(\frac{1}{T}\sum_{t=1}^{T} w^{(t)}\right)\right] - g(w^*) \\
&\leq \mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{T}\sum_{t=1}^{T} g(w^{(t)})\right] - g(w^*)
\end{aligned}$$

w^* does not depend on $v_1, \ldots, v_T$. Thus $g(w^*) = \mathbb{E}_{v_1,\ldots,v_T}\left[g(w^*)\right]$. Plugging it in the above inequality while using lemmas 2 and 3, we get:

$$\begin{aligned}
\mathbb{E}_{v_1,\ldots,v_T}\left[g(\bar{w})\right] - g(w^*) &\leq \mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{T}\sum_{t=1}^{T} g(w^{(t)})\right] - g(w^*) \\
&= \mathbb{E}_{v_1,\ldots,v_T}\left[\frac{1}{T}\sum_{t=1}^{T}\Big(g(w^{(t)}) - g(w^*)\Big)\right] \\
&\leq \mathbb{E}_{v_1,\ldots,v_T}\left[\sum_{t=1}^{T}\Big\langle w^{(t)} - w^*, \nabla g(w^{(t)})\Big\rangle\right] \\
&\leq \frac{B\rho}{\sqrt{T}}
\end{aligned}$$

□