We denote the synthesized label from greedy LAST as $y^*_{i_t}$. The other notation follows the main paper.

$$
\begin{aligned}
\left\|\boldsymbol{w}^{t+1}-\boldsymbol{w}^*\right\|^2 &= \left\|\boldsymbol{w}^t-\eta_t\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},y^*_{i_t}|\boldsymbol{w}^t)-\boldsymbol{w}^*\right\|^2\\
&= \left\|\boldsymbol{w}^t-\eta_t g(y^*_{i_t})\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t)-\boldsymbol{w}^*\right\|^2\\
&= \left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2-2\eta_t g(y^*_{i_t})\langle\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t),\boldsymbol{w}^t-\boldsymbol{w}^*\rangle+\eta_t^2(g(y^*_{i_t}))^2\left\|\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t)\right\|^2\\
① \quad &= \left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2+\min_{y_{i_t}}\left\{\eta_t^2(g(y_{i_t}))^2\left\|\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t)\right\|^2-2\eta_t g(y_{i_t})\langle\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t),\boldsymbol{w}^t-\boldsymbol{w}^*\rangle\right\}\\
② \quad &\leq \left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2+\eta_t^2c_1^2\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2\left\|\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t)\right\|^2-2\eta_t c_1\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|\langle\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t),\boldsymbol{w}^t-\boldsymbol{w}^*\rangle\\
③ \quad &\leq \left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2+\eta_t^2c_1^2L^2_{\max}\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2-2\eta_t c_1\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|\langle\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t),\boldsymbol{w}^t-\boldsymbol{w}^*\rangle\\
④ \quad &\leq \left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2+\eta_t^2c_1^2L^2_{\max}\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2-\eta_t c_1\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|^2\\
&\leq \cdots
\end{aligned}
$$

The final form is the same as Eq. (12) in the appendix and the rest of the proof follows from there.

- ① : According to the definition of greedy LAST, $y^*_{i_t}$ is the minimizer of the following optimization:

$$\min_{y_{i_t}}\left\{\eta_t^2(g(y_{i_t}))^2\left\|\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t)\right\|^2-2\eta_t g(y_{i_t})\langle\nabla_{\boldsymbol{w}^t}\ell(\boldsymbol{x}_{i_t},\tilde{y}_{i_t}|\boldsymbol{w}^t),\boldsymbol{w}^t-\boldsymbol{w}^*\rangle\right\}$$

- ② : Choose a y_{i_t} such that $g(y_{i_t})=c_1\left\|\boldsymbol{w}^t-\boldsymbol{w}^*\right\|$ and put it back to the minimization objective. By doing so, the resulting objective function becomes an upper bound of the original minimization objective, because any y_{i_t} will lead to a larger (or equal) objective value.
- ③ : Lipschitz continuous (see the proof of Theorem 1)
- ④ : Convexity and interpolation property (see the proof of Theorem 1)