from the lemma, weights $c_{t,i}$ is also symmetric, so this means actually when error $r_{t,i}$ gets further against zero, weights get smaller), but for least square regression, all weights are equal so there is no difference between outliers and inliers. Thus this algorithm achieves robustness since it decreases the importance for outliers.

3) Logistic Regression as ERM

Recall logistic regression: (for labels $Y \in \{-1, 1\}$)

$$f^*(\vec{x}) = \begin{cases} 1 & \text{if } \eta(x) \geq \frac{1}{2} \\ -1 & \text{otherwise.} \end{cases}$$

$\eta(\vec{x}) := Pr\{Y=1 \mid X=\vec{x}\}$ and we know $\eta(\vec{x}) = \dfrac{1}{1+\exp[-(w^T x+b)]}$

$$\Rightarrow Pr\{Y=-1 \mid X=\vec{x}\} = 1-\eta(\vec{x}) = \frac{1}{1+\exp(w^T x+b)}$$

$$\Rightarrow P(y \mid \vec{x}; \vec{\theta}) = \begin{cases} 1-\eta(\vec{x}) & \text{if } y=-1 \\ \eta(\vec{x}) & \text{if } y=1 \end{cases}$$

$$= \frac{1}{1+\exp[-y(w^T x+b)]}$$

Then the log-likelihood of $\vec{\theta}$ is defined to be.

$$\ell(\vec{\theta}) = \log L(\vec{\theta}) = -\sum_{i=1}^{n} \log(1+\exp[-y_i(w^T x_i+b)])$$

$$\left[= \log\left(\prod_{i=1}^{n} P(y_i \mid \vec{x}_i; \vec{\theta})\right)\right]$$

As for ERM with the logistic loss is $\frac{1}{n}\sum_{i=1}^{n} \log(1+\exp(-y(w^T x_i+b)))$, which is just proportional to the negative log likelihood for logistic regression