P8.5

(a) because the number of misclassified training points is the number of points with $z_i<0$

so $R_{emp} = \sum_{i=1}^{n} L_{01}(z_i)$ where $L_{01}(z_i) = \begin{cases} 0, & z_i \geq 0 \\ 1, & z_i < 0 \end{cases}$

(b) because the perception minimizes the error over the misclassified points

so $E(w) = \sum_{i \in \text{misclassified}} -z_i = \sum_{i=1}^{n} \begin{cases} 0, & z_i \geq 0 \\ -z_i, & z_i < 0 \end{cases} = \sum_{i=1}^{n} \max(0, -z_i)$

$\therefore L_p(z_i) = \max(0, -z_i)$

(c) $E(w) = \sum_{i=1}^{n} (w^T x_i - y_i)^2 = \sum_{i=1}^{n} \frac{1}{y_i^2} (y_i w^T x_i - y_i^2)^2 = \sum_{i=1}^{n} (y_i w^T x_i - 1)^2 \quad \because y_i^2 = 1$

$= \sum_{i=1}^{n} (z_i - 1)^2$

$\therefore L_{LSC}(z_i) = (z_i - 1)^2$

(d) we have $E(w) = -\sum_{i=1}^{n} \begin{cases} \log \sigma(w^T x_i), & y_i = 1 \\ \log \sigma(-w^T x_i), & y_i = -1 \end{cases}$

$= -\sum_{i=1}^{n} \log \sigma(y_i w^T x_i)$

$= \sum_{i=1}^{n} -\log \frac{1}{1+\exp(-y_i w^T x_i)}$

$= \sum_{i=1}^{n} \log(1+e^{-z_i})$

$\therefore L_{LR}(z_i) = \log(1+e^{-z_i}) \propto \frac{1}{\log(2)} \log(1+e^{-z_i})$