4. (a) $\vec{h} = \tanh(b + W\vec{x})$

(i)

$$\hat{t} = \sigma(b_4 + \theta^T \vec{h})$$

$$E(x, t) = -(t\log(\hat{t}) + (1-t)\log(1-\hat{t}))$$

Want to find $\nabla_\theta E(x,t)$, by chain rule, we know $= \frac{\partial \hat{t}}{\partial \theta} \cdot \frac{\partial E}{\partial \hat{t}}$

First, compute $\frac{\partial E}{\partial \hat{t}}$

$$\Rightarrow \frac{\partial}{\partial t} E(x,t) = \frac{\partial}{\partial t} -(t\log(\hat{t}) + (1-t)\log(1-\hat{t}))$$

$$= \frac{-\partial}{\partial t} t\log(\hat{t}) - \frac{\partial}{\partial t}(1-t)(\log(1-\hat{t})) = \frac{1-t}{1-\hat{t}} - \frac{t}{\hat{t}} = \frac{\hat{t}-t}{\hat{t}(1-t)}$$

Next, compute $\frac{\partial \hat{t}}{\partial \theta} = \frac{\partial}{\partial \theta} \sigma(b_4 + \theta^T \vec{h}) = \sigma(b_4 + \theta^T \vec{h})(1 - \sigma(b_4 + \theta^T \vec{h})\vec{h}$

$$= \hat{t}(1-\hat{t})\vec{h}$$

Multiplying, we get

$$\nabla_\theta E(x,t) = \cancel{\hat{t}(1-\hat{t})}\vec{h} \cdot \frac{\hat{t}-t}{\cancel{\hat{t}(1-\hat{t})}} = \boxed{\vec{h}(\hat{t}-t)}$$

(ii) Same procedure as (a), except we want to compute $\frac{\partial \hat{t}}{\partial b_4} \cdot \frac{\partial E}{\partial \hat{t}}$

$$\frac{\partial \hat{t}}{\partial b_4} = \frac{\partial}{\partial b_4} \sigma(b_4 + \theta^T \vec{h}) = \sigma(b_4 + \theta^T \vec{h})(1 - \sigma(b_4 + \theta^T \vec{h})$$

$$= (\hat{t})(1-\hat{t})$$

$$\Rightarrow \frac{\partial}{\partial b_4} E(x,t) = \cancel{(\hat{t})(1-\hat{t})} \cdot \frac{\hat{t}-t}{\cancel{(\hat{t})(1-\hat{t})}} = \boxed{\hat{t}-t}$$

$\boxtimes$

4 (b) Want to find

(i)

$$\nabla_W E(\vec{x},t), \text{ by chain rule, } = \frac{\partial E}{\partial h} \cdot \frac{\partial h}{\partial W}$$

First, compute

$$\frac{\partial E}{\partial h} = \frac{\partial \hat{t}}{\partial h}\left[\frac{\partial}{\partial \hat{t}} E(\vec{x},t)\right] = \frac{\partial}{\partial h}\left[\sigma(b_q + \Theta^T \vec{h})\right]\left[\frac{\hat{t}-t}{\hat{t}(1-\hat{t})}\right]$$

$$= \sigma(b_q + \Theta^T \vec{h})(1-\sigma(b_q + \Theta^T \vec{h}))\,\Theta^T\left[\frac{\hat{t}-t}{\hat{t}(1-\hat{t})}\right]$$

$$= \cancel{(\hat{t})(1-\hat{t})}\,\Theta^T\left[\frac{\hat{t}-t}{\cancel{\hat{t}(1-\hat{t})}}\right] = \Theta^T(\hat{t}-t)$$

Next, compute

$$\frac{\partial h}{\partial W} = \frac{\partial}{\partial W}\tanh(\vec{b}+Wx) = \left[1-\tanh^2(\vec{b}+Wx)\right]\vec{x}$$

$$= \left[1-h^2\right]\vec{x}$$

Thus, $\nabla_W E(\vec{x},t) = \Theta^T(\hat{t}-t)(1-\vec{h}^2)\vec{x}$

(ii) Similar to (i), except we want

$$\nabla_b E(\vec{x},t) = \cancel{\frac{\partial E}{\partial h}\frac{\partial b}{\partial W}} \quad \frac{\partial E}{\partial h}\cdot\frac{\partial h}{\partial b}$$

Thus,

$$\cancel{\frac{\partial E}{\partial b}\frac{\partial b}{\partial W}} = \frac{\partial h}{\partial b} = \frac{\partial}{\partial b}\tanh(\vec{b}+Wx) = 1-\tanh^2(\vec{b}+W\vec{x})$$

$$= 1-h^2.$$

Thus, $\nabla_b E(\vec{x},t) = \Theta^T(\hat{t}-t)(1-\vec{h}^2)$ $\boxtimes$