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

Find gradients.

i) $\nabla_\theta E(x,t)$: $\frac{\partial E}{\partial \theta_i} = -\left(\frac{t}{\hat{t}} - \frac{1-t}{1-\hat{t}}\right)\frac{d\hat{t}}{d\theta_i}$

$$\frac{d\hat{t}}{d\theta_i} = \frac{h_i^{(2)} e^{-(b_4+\theta^T h^{(2)})}}{(1+e^{-(b_4+\theta^T h^{(2)})})^2} = -h_i^{(2)}(1-\hat{t})\hat{t}$$

So $$\nabla_\theta E(x,t) = -\left(\frac{t}{\hat{t}} - \frac{1-t}{1-\hat{t}}\right)(1-\hat{t})\hat{t}\,h^{(2)}$$

$$= -(t(1-\hat{t}) - \hat{t}(1-t))\,h^{(2)}$$

$$= (\hat{t}-t)\,h^{(2)}$$

and $$\frac{\partial E}{\partial b_4} = -\left(\frac{t}{\hat{t}} - \frac{1-t}{1-\hat{t}}\right)\frac{\partial \hat{t}}{\partial b_4}$$

$$= -\left(\frac{t}{\hat{t}} - \frac{1-t}{1-\hat{t}}\right)\frac{-e^{-(b_4+\theta^T h^{(2)})}}{(1+e^{-(b_4+\theta^T h^{(2)})})^2}$$

$$= -\left(\frac{t}{\hat{t}} - \frac{1-t}{1-\hat{t}}\right)(1-\hat{t})\hat{t}$$

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

ii) Find

$\nabla_{w_1} E$, $\nabla_{w_2} E$, $\nabla_{b_1} E$, and $\nabla_{b_2} E$.

$$\frac{dE}{dh_i^{(2)}} = -\left(\frac{t}{\hat{t}} - \frac{1-t}{1-\hat{t}}\right)(1-\hat{t})\hat{t}\,\theta_i = (\hat{t}-t)\theta_i$$

$$\frac{dh_i^{(2)}}{db_i^{(2)}} = 1-\tanh^2(b_2+W^{(2)}h^{(1)}) = 1-(h_i^{(2)})^2,$$ so

$$\frac{dE}{db_i^{(2)}} = (\hat{t}-t)(1-(h_i^{(2)})^2)\theta_i$$

and

$$\frac{dE}{dw_i^{(2)}} = (\hat{t}-t)(1-(h_i^{(2)})^2)\theta_i h_i^{(1)} .$$

while $$\frac{dE}{dh_i^{(1)}} = (\hat{t}-t)(1-(h_i^{(2)})^2)W_i^{(2)}\theta_i$$

and $$\frac{dh_i^{(1)}}{db_i^{(1)}} = (1-(h_i^{(1)})^2)$$ while $$\frac{dh_i}{dW_i^{(1)}} = X_i(1-(h_i^{(1)})^2)$$

so $$\frac{dE}{db_i^{(1)}} = (\hat{t}-t)(1-(h_i^{(1)})^2)(1-(h_i^{(2)})^2) W_i^{(2)}\theta_i$$

and $$\frac{dE}{dW_i^{(1)}} = (\hat{t}-t)(1-(h_i^{(1)})^2)(1-(h_i^{(2)})^2) W_i^{(2)}\theta_i x_i$$