1. a) (i) $\nabla_{w_y} \log p(y^t | x^t)$

$$\nabla_{w_y} \log p(y^t | x^t) = \nabla_{w_y} \left(-\log Z_{x^t} + \sum_{s=1}^{m} \langle w_{y_s^t}, x_s^t \rangle + \sum_{s=1}^{m-1} T_{y_s^t, y_{s+1}^t} \right) \quad -(5)$$

$$= \nabla_{w_y} \left(-\log Z_{x^t} + \sum_{s=1}^{m} \langle w_{y_s^t}, x_s^t \rangle \right) \quad -(6)$$

Taking gradient of second term:-

$$\nabla_{w_y} \sum_{s=1}^{m} \langle w_{y_s^t}, x_s^t \rangle = \sum_{s=1}^{m} \nabla_{w_y} (w_{y_s^t}^T x_s^t) \quad -(7)$$

$$= \sum_{s=1}^{m} [y_s^t = y] \, x_s^t \quad -(8)$$

Taking gradient of first term:-

$$-\nabla_{w_y} \log Z_{x^t} = -\frac{1}{Z_{x^t}} \sum_{y \in \mathcal{Y}^m} \exp\left(\sum_{s=1}^{m} \langle w_{y_s}, x_s^t \rangle + \sum_{s=1}^{m-1} T_{y_s, y_{s+1}} \right) \nabla_{w_y} \sum_{s=1}^{m} \langle w_{y_s}, x_s^t \rangle \quad -(9)$$

$$= -\sum_{y \in \mathcal{Y}^m} p(y | x^t) \sum_{s=1}^{m} [y_s = y] \, x_s^t \quad -(10)$$

$$= -\sum_{s=1}^{m} p(y_s = y | x^t) \, x_s^t \quad -(11)$$

Therefore, we got:

$$\nabla_{w_y} \log p(y^t | x^t) = \sum_{s=1}^{m} \left([y_s^t = y] - p(y_s = y | x^t) \right) x_s^t \quad -(12)$$

(ii) $\nabla_{T_{ij}} \log p(y^t | x^t)$

$$\nabla_{T_{ij}} \log p(y^t | x^t) = \nabla_{T_{ij}} \left(-\log Z_{x^t} + \sum_{s=1}^{m} \langle w_{y_s^t}, x_s^t \rangle + \sum_{s=1}^{m-1} T_{y_s^t, y_{s+1}^t} \right) \quad -(13)$$

$$= \nabla_{T_{ij}} \left(-\log Z_{x^t} + \sum_{s=1}^{m-1} T_{y_s^t, y_{s+1}^t} \right) \quad -(14)$$