9.3.

$$J(B) = -\sum_{i=1}^{N}\left[\sum_{g=1}^{r} p(g|x_i)\, s_g(x_i) - \log\left(\sum_{h=1}^{r} \exp\left(s_h(x_i)\right)\right)\right]$$

$$\nabla_g J(B) = \begin{pmatrix} \frac{\partial J(B)}{\partial B_0^{(g)}} \\ \vdots \\ \frac{\partial J(B)}{\partial B_0^{(g)}} \end{pmatrix}$$

$$\nabla_g J(B) = -\sum_{i=1}^{N}\left[\nabla_g p(g|x_i)\, s_g(x_i) - \nabla_g \log\left(\sum_{h=1}^{r} \exp(s_h(x_i))\right)\right]$$

$$= -\sum_{i=1}^{N}\left[p(g|x_i)\nabla_g s_g(x_i) - \frac{\nabla_g \exp(s_g(x_i))}{\sum_{h=1}^{r} \exp(s_h(x_i))}\right]$$

$$= -\sum_{i=1}^{N}\left[p(g|x_i)\begin{pmatrix}1\\ x_i\end{pmatrix} - \frac{\exp(s_g(x_i))}{\sum_{i=1} \exp(s_h(x_i))}\begin{pmatrix}1\\ x_i\end{pmatrix}\right]$$

$$= -\sum_{i=1}^{N}\left[p(g|x_i)\begin{pmatrix}1\\ x_i\end{pmatrix} - p_B(g|x_i)\begin{pmatrix}1\\ x_i\end{pmatrix}\right]$$

$$= \sum_{i=1}^{N}\left[p_B(g|x_i) - p(g|x_i)\right]\begin{pmatrix}1\\ x_i\end{pmatrix}$$