Let

$$L(W) = \frac{1}{2} \sum_{i=1}^{K} \| b^{(i)} - W a^{(i)} \|_2^2$$

Recall that for a matrix $X \in \mathbb{R}^{m \times n}$

$$\|X\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij})^2$$

$$= \sum_{j=1}^{n} \|X(:, j)\|_2^2$$

↳ 2-norm of $j^{\underline{th}}$ column of X

So if we define the following matrices

$$B = \begin{bmatrix} | & | & & | \\ b^{(1)} & b^{(2)} & \cdots & b^{(K)} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{m \times K}$$

$$A = \begin{bmatrix} | & | & & | \\ a^{(1)} & a^{(2)} & \cdots & a^{(K)} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times K}$$

Then,

$$L(W) = \frac{1}{2} \| B - WA \|_F^2$$

Using the trace definition of frobenius norm,

$$L(W) = \frac{1}{2} Tr\left[(B - WA)^T (B - WA) \right]$$

$$= \frac{1}{2} Tr\left[B^T B - B^T WA - A^T W^T B + A^T W^T WA \right]$$

Dropping the terms that has no W dependence and observing

$$Tr\left[B^T W A\right] = Tr\left[(B^T WA)^T \right]$$

$$= Tr\left[A^T W^T B \right]$$

We have

$$L(W) = -Tr\left[A^T W^T B\right] + \frac{1}{2} Tr\left[A^T W^T WA\right]$$

Now by cyclic property of Trace

$$Tr[A^T W^T W A]$$

$$= Tr[W^T W A A^T]$$

Hence,

$$L(W) = -Tr[A^T W^T B] + \frac{1}{2} Tr[W^T W A A^T]$$

Now,

$$\nabla_W Tr[A^T W^T B] = BA^T$$ [(102) in matrix cookbook]

$$\nabla_W Tr[W^T W A A^T] = 2WAA^T$$ [(113) in matrix cookbook]

Hence,

$$\nabla_W L(W) = -BA^T + WAA^T$$