Proof Sketch of Theorem 4.8 without fixing classifiers. Let $S := \sum_{k=1}^{K} f_k n_k$, and $\overline{\mathbf{w}}^{(k)} := \frac{1}{f_k}\sum_{j=1}^{f_k} \mathbf{w}_j^{(k)^\top}$

$$
\begin{aligned}
& \frac{1}{2S}\|\mathbf{W}\mathbf{H}-\mathbf{Y}\|_F^2 + \frac{\lambda_H}{2}\|\mathbf{H}\|_F^2 + \frac{\lambda_{\mathbf{W}_0}}{2}\|\mathbf{W_0}\|_F^2 && (1)\\
& \overset{(a)}{\geq} \frac{1}{2S}\sum_{k=1}^{K}\sum_{i=j}^{f_k} n_k \frac{1}{n_k}\sum_{i=1}^{n_k}\left(\mathbf{w}_j^{(k)^\top}\mathbf{h}_{k,i}-1\right)^2 + \frac{\lambda_H}{2}\sum_{k=1}^{K} n_k \frac{1}{n_k}\sum_{i=1}^{n_k}\|\mathbf{h}_{k,i}\|_2^2 + \frac{\lambda_{\mathbf{W}_0}}{2}\sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2 && (2)\\
& \overset{(b)}{\geq} \frac{1}{2S}\sum_{k=1}^{K}\sum_{j=1}^{f_k} n_k \left(\mathbf{w}_j^{(k)^\top}\frac{1}{n_k}\sum_{i=1}^{n_k}\mathbf{h}_{k,i}-1\right)^2 + \frac{\lambda_H}{2}\sum_{k=1}^{K} n_k \left\|\frac{1}{n_k}\sum_{i=1}^{n_k}\mathbf{h}_{k,i}\right\|_2^2 + + \frac{\lambda_{\mathbf{W}_0}}{2}\sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2 && (3)\\
& \overset{(c)}{\geq} \frac{1}{2S}\sum_{k=1}^{K} n_k f_k \left(\overline{\mathbf{w}}^{(k)^\top}\mathbf{h_k}-1\right)^2 + \frac{\lambda_H}{2}\sum_{k=1}^{K} n_k \|\mathbf{h}_k\|_2^2 + \frac{\lambda_{\mathbf{W}_0}}{2}\sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2 && (4)\\
& \geq \frac{1}{2S}\min_{\mathbf{H},\mathbf{W}}\sum_{k=1}^{K} n_k \left[f_k\left(\overline{\mathbf{w}}^{(k)^\top}\mathbf{h_k}-1\right)^2 + \lambda_H S\|\mathbf{h}_k\|_2^2 + \frac{\lambda_{\mathbf{W}_0} S}{n_k}\|\mathbf{w}_k\|_2^2\right] && (5)\\
& \overset{(d)}{=} \frac{1}{2S}\sum_{k=1}^{K} n_k \min_{\mathbf{h}_k,\mathbf{w}_k}\left[f_k\left(\overline{\mathbf{w}}^{(k)^\top}\mathbf{h_k}-1\right)^2 + \lambda_H S\|\mathbf{h}_k\|_2^2 + \frac{\lambda_{\mathbf{W}_0} S}{n_k}\|\mathbf{w}_k\|_2^2\right] && (6)\\
& \overset{(e)}{\geq} \frac{1}{2S}\sum_{k=1}^{K} n_k \min_{\mathbf{h}_k,\mathbf{w}_k}\left[f_k\left(\overline{\mathbf{w}}^{(k)^\top}\mathbf{h_k}-1\right)^2 + 2S\sqrt{\frac{\lambda_H\lambda_{\mathbf{W}_0}}{n_k}}\|\mathbf{h}_k\|\|\mathbf{w}_k\|\right] && (7)
\end{aligned}
$$

We observe that the equality of (e) holds only when $\lambda_{\mathbf{W}_0}||\mathbf{w}_k||^2 = n_k\lambda_{\mathbf{H}}||\mathbf{h}_k||^2$ by Young's Inequality. □