Hyperparameter optimization with Lagrange multipliers

Henrik Lund Mortensen

October 3, 2017

Problem:
Minimize $E_{\text{val}}(\boldsymbol{w})$ with respect to σ, where $\boldsymbol{w} = \left(\boldsymbol{K} + \gamma\frac{N}{2}\right)^{-1}\boldsymbol{Y}$

Instead of $E_{\text{val}}(\tilde{\boldsymbol{w}})$ we minimize the Lagrangian

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{p}, \sigma) = E_{\text{val}}(\boldsymbol{w}) + \boldsymbol{p} \cdot \left(\frac{2}{N}\left(\boldsymbol{K}\boldsymbol{w} - \boldsymbol{Y}\right) + \gamma\boldsymbol{w}\right) \tag{1}$$

The Lagrangian is minimized by setting all partial derivative to zero

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = 0 \ , \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{p}} = 0 \ , \quad \frac{\partial \mathcal{L}}{\partial \sigma} = 0 \tag{2}$$

The derivative with respect to the multiplier gives

$$0 = \frac{\partial \mathcal{L}}{\partial \boldsymbol{p}} = \frac{2}{N}\left(\boldsymbol{K}\boldsymbol{w} - \boldsymbol{Y}\right) + \gamma\boldsymbol{w} \tag{3}$$

$$\boldsymbol{w} = \left(\boldsymbol{K} + \gamma\frac{N}{2}\right)^{-1}\boldsymbol{Y}, \tag{4}$$

which is just the condition on $\boldsymbol{w}$. The derivative of $\mathcal{L}$ with respect to $\boldsymbol{w}$ ives

$$0 = \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \frac{\partial E_{\text{val}}(\boldsymbol{w})}{\partial \boldsymbol{w}} + \boldsymbol{p} \cdot \left(\gamma + \frac{2}{N}\boldsymbol{K}\right) \tag{5}$$

$$\boldsymbol{p} = -\frac{\partial E_{\text{val}}(\boldsymbol{w})}{\partial \boldsymbol{w}}\left(\gamma + \frac{2}{N}\boldsymbol{K}\right)^{-1} \tag{6}$$