A Proof of Proposition 1

Proposition 1. Define $I_L(W) := E_{x \sim L}[I(x, W)]$ and $I_X(W) := E_{x \sim X}[I(x, W)]$, the expected loss error of the semi-supervised estimate $E[G(W^L, X) - G(W^*, X)]$ with respect to $B$ ground truths sampled from $L$ is upper bounded by $O\left( (1 + \frac{\lambda B}{N}) \tr \left( (I_X(W^*) + \frac{\lambda B}{N} I_L(W^*))^{-1} I_X(W^*) \right) \right)$.

Proof. We follow the notations of Chaudhuri et al. [1]. We denote $\psi_i(W) := U(W^*) - \lambda B \log p(y_i | x_i, W), \forall x_i \in L$, where $y_i$ is sampled from $p(y_i | x_i, W^*)$. When $E[xx^\top]$ exists and is positive definite, $\psi_i(W)$ is smooth and strong convex. We denote $P(W) := E[\psi_i(W)]$ and $Q(W) := G_X(W)$, and the latter is the expected loss when the distribution of the ground truths for all tasks are observed. We also have $\nabla Q(W^*) = 0$.

The Hessian of the loss on one verification sample $x$ is

$$\frac{\partial^2 G(W, x)}{\partial W^2} = -N \cdot \frac{\partial^2 U(W)}{\partial W^2} - \lambda B \cdot \frac{\partial^2 \log p(y | x, W)}{\partial W^2}$$

$$= N \cdot E_{x \sim X}[I(x, W)] + \lambda B \cdot I(x, W)$$

$$= N \cdot I_X(W) + \lambda B \cdot I(x, W)$$.

Then we directly apply the Lemma 1 of Chaudhuri et al. [1] on $G_X(W)$, we have that

$$E[G(W^L, X) - G(W^*, X)] = O\left( (1 + \frac{\lambda B}{N}) \tr \left( (I_X(W^*) + \frac{\lambda B}{N} I_L(W^*))^{-1} I_X(W^*) \right) \right)$$.

Here we ignore all the constants and small quantities, since we only care about the relationship between the expected loss error and the verification subset.

Reference