
Algorithm 1 Minimax SGD

- 1: initialize all hyper-parameters, and initialize the primal variable θ and the dual variables \mathbf{w}_1 and \mathbf{w}_2 .
 - 2: $g(\theta) = [P_\alpha(\mathbf{v}|\mathbf{w}), \log \frac{P(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)}]$.
 - 3: $\mathbf{y} = \text{Sigmoid}(\mathbf{e}_1 \text{Sigmoid}(\theta \mathbf{e}_2 + \mathbf{b}_2) + \mathbf{b}_1)$.
 - 4: **update the primal variable:**
 - 5: $\Delta\theta = \frac{\partial \mathbf{y}_1}{\partial \theta} g_1(\theta) + \frac{\partial \mathbf{y}_2}{\partial \theta} g_2(\theta) + \frac{\partial g_1(\theta)}{\partial \theta} \mathbf{y}_1 + \frac{\partial g_2(\theta)}{\partial \theta} \mathbf{y}_2 - \frac{\partial f^*(\mathbf{y})}{\partial \theta}$
 - 6: $\theta = \theta - \alpha \Delta\theta$.
 - 7: **update the dual variable:**
 - 8: $\mathbf{e}_1 = \mathbf{e}_1 + \beta \left(g(\theta) \frac{\partial \mathbf{y}}{\partial \mathbf{e}_1} - \nabla f^*(\mathbf{y}) \frac{\partial \mathbf{y}}{\partial \mathbf{e}_1} \right)$.
 - 9: $\mathbf{e}_2 = \mathbf{e}_2 + \beta \left(g(\theta) \frac{\partial \mathbf{y}}{\partial \mathbf{e}_2} - \nabla f^*(\mathbf{y}) \frac{\partial \mathbf{y}}{\partial \mathbf{e}_2} \right)$.
 - 10: $\mathbf{b}_1 = \mathbf{b}_1 + \beta \left(g(\theta) \frac{\partial \mathbf{y}}{\partial \mathbf{b}_1} - \nabla f^*(\mathbf{y}) \frac{\partial \mathbf{y}}{\partial \mathbf{b}_1} \right)$.
 - 11: $\mathbf{b}_2 = \mathbf{b}_2 + \beta \left(g(\theta) \frac{\partial \mathbf{y}}{\partial \mathbf{b}_2} - \nabla f^*(\mathbf{y}) \frac{\partial \mathbf{y}}{\partial \mathbf{b}_2} \right)$.
-

$$\begin{aligned} L(\theta) &= \int \log(E_{p_\alpha(\mathbf{w})}(p_\alpha(\mathbf{v}|\mathbf{w}))) q(\mathbf{v}|\theta) d\mathbf{v} + \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta) d\mathbf{v} \\ &= \int \log \left(\int p_\alpha(\mathbf{v}|\mathbf{w}) p_\alpha(\mathbf{w}) d\mathbf{w} \right) q(\mathbf{v}|\theta) d\mathbf{v} + \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta) d\mathbf{v} \\ &= \int \log \left(\int \frac{1}{|K_{nn}|^{1/2}} e^{-\frac{1}{2}(\mu+L\epsilon)K_{nn}^{-1}(\mu+L\epsilon)} e^{-\frac{1}{2}(\log \mathbf{w}-\mu_0)(\mathbf{e}_0^2)^{-1}(\log \mathbf{w}-\mu_0)} d\mathbf{w} \right) \frac{1}{|LL^T|^{1/2}} e^{-\frac{1}{2}\epsilon^T \epsilon} d\mathbf{v} \\ &+ \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta) d\mathbf{v} \\ &= \int \log \left(\int \frac{1}{|K_{nn}|^{1/2}} e^{-\frac{1}{2}(\mu+L\epsilon)K_{nn}^{-1}(\mu+L\epsilon)} e^{-\frac{1}{2}(\log \mathbf{w}-\mu_0)(\mathbf{e}_0^2)^{-1}(\log \mathbf{w}-\mu_0)} d\mathbf{w} \right) \frac{1}{|LL^T|^{1/2}} e^{-\frac{1}{2}\epsilon^T \epsilon} d\mathbf{v} \\ &+ \int (\log p_\alpha(\mathbf{D}|\mathbf{v}) - \log q(\mathbf{v}|\theta)) q(\mathbf{v}|\theta) d\mathbf{v} \\ &= \int \log \left(\int \frac{1}{|K_{nn}|} e^{-(\mu+L\epsilon)K_{nn}^{-1}(\mu+L\epsilon)} e^{-(\log \mathbf{w}-\mu_0)(\mathbf{e}_0^2)^{-1}(\log \mathbf{w}-\mu_0)} d\mathbf{w} \right) \frac{1}{|LL^T|} e^{-\epsilon^T \epsilon} d\mathbf{v} \\ &+ \int \left(\sum_{i=1}^{n-n_{test}} (-\log(1 + e^{-r_i(\mu_i+L_i\epsilon)}) + \log |LL^T| + \epsilon^T \epsilon) \right) \frac{1}{|LL^T|} e^{-\epsilon^T \epsilon} d\mathbf{v} \\ &= \int \log \left(\int \frac{1}{|K_{nn}|} e^{-(\mu+L\epsilon)K_{nn}^{-1}(\mu+L\epsilon)} e^{-(\log \mathbf{w}-\mu_0)(\mathbf{e}_0^2)^{-1}(\log \mathbf{w}-\mu_0)} d\mathbf{w} \right) \frac{1}{|LL^T|} e^{-\epsilon^T \epsilon} d(\mu + L\epsilon) \\ &+ \int \left(\sum_{i=1}^{n-n_{test}} (-\log(1 + e^{-r_i(\mu_i+L_i\epsilon)}) + \log |LL^T| + \epsilon^T \epsilon) \right) \frac{1}{|LL^T|} e^{-\epsilon^T \epsilon} d(\mu + L\epsilon) \end{aligned}$$

1 References