Algorithm 1 Minimax SGD

initialize all hyper-parameters, and initialize the primal variable θ and the dual variables $\mathbf{w}_1$ and $\mathbf{w}_2$.
for $t = 1 : T$ **do**
  sample ϵ to compute $\mathbf{v} = \mu + L\epsilon$.
  sample $\mathbf{w}$.
  $g(\theta) = [P_\alpha(\mathbf{v}|\mathbf{w}), \log \frac{P(D|\mathbf{v})}{q(\mathbf{v}|\theta)}]$.
  $\mathbf{y} = \mathbf{e}_1 Sigmoid(\mathbf{e}_2 \epsilon + \mathbf{b}_2) + \mathbf{b}_1$.
  update the primal variable:
  $\Delta\theta = \frac{\partial g_1(\theta)}{\partial \theta}\mathbf{y}_1 + \frac{\partial g_2(\theta)}{\partial \theta}\mathbf{y}_2$.
  $\theta = \theta - \alpha\Delta\theta$.
  update the dual variable:
  $\mathbf{e}_1 = \mathbf{e}_1 + \beta\left(g(\theta)\frac{\partial \mathbf{y}}{\partial \mathbf{e}_1} - \nabla f^*(\mathbf{y})\frac{\partial \mathbf{y}}{\partial \mathbf{e}_1}\right)$.
  $\mathbf{e}_2 = \mathbf{e}_2 + \beta\left(g(\theta)\frac{\partial \mathbf{y}}{\partial \mathbf{e}_2} - \nabla f^*(\mathbf{y})\frac{\partial \mathbf{y}}{\partial \mathbf{e}_2}\right)$.
  $\mathbf{b}_1 = \mathbf{b}_1 + \beta\left(g(\theta)\frac{\partial \mathbf{y}}{\partial \mathbf{b}_1} - \nabla f^*(\mathbf{y})\frac{\partial \mathbf{y}}{\partial \mathbf{b}_1}\right)$.
  $\mathbf{b}_2 = \mathbf{b}_2 + \beta\left(g(\theta)\frac{\partial \mathbf{y}}{\partial \mathbf{b}_2} - \nabla f^*(\mathbf{y})\frac{\partial \mathbf{y}}{\partial \mathbf{b}_2}\right)$.

$$
\begin{aligned}
L(\theta) &= \int \log(E_{p_\alpha(\mathbf{w})}(p_\alpha(\mathbf{v}|\mathbf{w})))q(\mathbf{v}|\theta)\mathrm{d}\mathbf{v} + \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta)\mathrm{d}\mathbf{v} \\
&= \int \log\left(\int p_\alpha(\mathbf{v}|\mathbf{w})p_\alpha(\mathbf{w})\mathrm{d}\mathbf{w}\right) q(\mathbf{v}|\theta)\mathrm{d}\mathbf{v} + \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta)\mathrm{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)}\mathrm{d}\mathbf{w}\right) \frac{1}{|LL^T|^{1/2}} e^{-\frac{1}{2}\epsilon^T\epsilon}\mathrm{d}\mathbf{v} \\
&+ \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta)\mathrm{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)}\mathrm{d}\mathbf{w}\right) \frac{1}{|LL^T|^{1/2}} e^{-\frac{1}{2}\epsilon^T\epsilon}\mathrm{d}\mathbf{v} \\
&+ \int (\log p_\alpha(\mathbf{D}|\mathbf{v}) - \log q(\mathbf{v}|\theta))q(\mathbf{v}|\theta)\mathrm{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)}\mathrm{d}\mathbf{w}\right) \frac{1}{|LL^T|} e^{-\epsilon^T\epsilon}\mathrm{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}\mathrm{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)}\mathrm{d}\mathbf{w}\right) \frac{1}{|LL^T|} e^{-\epsilon^T\epsilon}\mathrm{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}\mathrm{d}(\mu+L\epsilon)
\end{aligned}
$$

References