Algorithm 1 Minimax SGD

```
initialize all hyper-parameters, and initialize the primal variable θ and the dual variables w_1 and w_2.
for t = 1 : T do
    sample ϵ to compute v = μ + Lϵ.
    sample w to compute ARD kernel matrix.
    g(θ) = [P_α(v|w), log P(D|v)/q(v|θ)].
    y = e_1 Sigmoid(e_2 ϵ + b_2) + b_1.
    update the primal variable:
    Δθ = ∂g_1(θ)/∂θ y − ∂g_2(θ)/∂θ.
    θ = θ − αΔθ.
    update the dual variable:
    e_1 = e_1 + β (g_1(θ) + 1/y) ∂y/∂e_1.
    e_2 = e_2 + β (g_1(θ) + 1/y) ∂y/∂e_2.
    b_1 = b_1 + β (g_1(θ) + 1/y) ∂y/∂b_1.
    b_2 = b_2 + β (g(θ) + 1/y) ∂y/∂b_2.
```

$$
\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} \\
= \quad & \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} \\
= \quad & \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} \\
+ \quad & \int \log \frac{p_\alpha(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} q(\mathbf{v}|\theta)\mathrm{d}\mathbf{v} \\
= \quad & \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} \\
+ \quad & \int (\log p_\alpha(\mathbf{D}|\mathbf{v}) - \log q(\mathbf{v}|\theta)) q(\mathbf{v}|\theta)\mathrm{d}\mathbf{v} \\
= \quad & \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} \\
+ \quad & \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} \\
= \quad & \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) \\
+ \quad & \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}
$$

$$
g2 = \log \frac{p(\mathbf{D}|\mathbf{v})}{q(\mathbf{v}|\theta)} = \sum_i (-\log(1+\exp(r(i)(\mu_i + L_i\epsilon)))) + \log(|LL^T|))
$$

$$
\frac{\partial q(\mathbf{v}|\theta)}{\partial L} = (L^T)^{-1}
$$

References