**Supplementary Note B**

**Algorithms for implementing stochastic gradient decent**

Algorithm is a modified version of Algorithms in (4).

Let $l$ be the number of steps to apply to the discriminator.

**for** number of training iterations **do**

 **for** $l$ steps **do**

1. Sampling minibatch of $m$ noise vectors $\left\{z^{(1)},…,z^{(m)}\right\}$ from prior $P\_{g}(z)$.
2. Sampling minibatch of $m$ samples of feature, observed outcomes, treatment and treatment assignment indicator data $\left\{X^{(i)}, T^{(i)}, M^{(i)}, y\_{f}^{(i)}, i=1,…,m\right\}$from data generating distributions.
3. Update the parameters in discriminator by descending its stochastic gradient:

$-∇\_{θ\_{d}}\hat{V}(D\_{G},G,θ\_{d})$ . (B1)

 **end if**

1. Sampling minibatch of $m$ noise vectors $\left\{z^{(1)},…,z^{(m)}\right\}$ from prior $P\_{g}(z)$.
2. Update the parameters in generator by descending its stochastic gradient:

$∇\_{θ\_{g}}\hat{V}\left(D\_{G},G\left(θ\_{g}\right)\right)+λl(G\left(θ\_{g}\right))$. (B2)

**end if**