Supplementary Materials for "Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift"

A. Details of formulas in Section 3.2 (paper)

In this section, we list the detailed deductions of formulas especially in Section 3.2 (paper). Firstly, $Var^{Train}(X)$ is expanded as:

$$\begin{aligned}
&Var^{Train}(X) = Cov(\sum_{i=1}^{d} w_i a_i \frac{1}{p} x_i, \sum_{i=1}^{d} w_i a_i \frac{1}{p} x_i) \\
&= \frac{1}{p^2} \sum_{i=1}^{d} (w_i)^2 Var(a_i x_i) \\
&+ \frac{1}{p^2} \sum_{i=1}^{d} \sum_{j \neq i}^{d} \rho_{i,j}^{ax} w_i w_j \sqrt{Var(a_i x_i)} \sqrt{Var(a_j x_j)} \\
&= (\frac{1}{p}(c^2+v) - c^2)(\sum_{i=1}^{d} w_i^2 + \rho^{ax} \sum_{i=1}^{d} \sum_{j \neq i}^{d} w_i w_j),
\end{aligned} \tag{1}$$

and $Var^{Test}(X)$ is obtained:

$$\begin{aligned}
&Var^{Test}(X) = Var(\sum_{i=1}^{d} w_i x_i) \\
&= Cov(\sum_{i=1}^{d} w_i x_i, \sum_{i=1}^{d} w_i x_i) = \sum_{i=1}^{d} w_i^2 Var(x_i) \\
&+ \sum_{i=1}^{d} \sum_{j \neq i}^{d} \rho_{i,j}^{x} w_i w_j \sqrt{Var(x_i)} \sqrt{Var(x_j)} \\
&= v(\sum_{i=1}^{d} w_i^2 + \rho^{x} \sum_{i=1}^{d} \sum_{j \neq i}^{d} w_i w_j).
\end{aligned} \tag{2}$$

Further we can get the relation between ρ^{ax} and ρ^{x}:

$$\begin{aligned}
&\rho^{ax} = \rho_{i,j}^{ax} = \frac{Cov(a_i x_i, a_j x_j)}{\sqrt{Var(a_i x_i)} \sqrt{Var(a_j x_j)}} \\
&= \frac{p^2 Cov(x_i, x_j)}{\frac{p(c^2+v) - p^2 c^2}{v} \sqrt{Var(x_i)} \sqrt{Var(x_j)}} \\
&= \frac{v}{\frac{1}{p}(c^2+v) - c^2} \rho_{i,j}^{x} = \frac{v}{\frac{1}{p}(c^2+v) - c^2} \rho^{x}.
\end{aligned} \tag{3}$$

According to the above equations, the variance shift for case (b) can be written as:

$$\begin{aligned}
&\triangle(p,d) = \frac{Var^{Test}(X)}{Var^{Train}(X)} \\
&= \frac{v(\sum_{i=1}^{d} w_i^2 + \rho^{x} \sum_{i=1}^{d} \sum_{j \neq i}^{d} w_i w_j)}{(\frac{1}{p}(c^2+v) - c^2)(\sum_{i=1}^{d} w_i^2 + \rho^{ax} \sum_{i=1}^{d} \sum_{j \neq i}^{d} w_i w_j)} \\
&= \frac{v \sum_{i=1}^{d} w_i^2 + v \rho^{x} \sum_{i=1}^{d} \sum_{j \neq i}^{d} w_i w_j}{(\frac{1}{p}(c^2+v) - c^2) \sum_{i=1}^{d} w_i^2 + v \rho^{x} \sum_{i=1}^{d} \sum_{j \neq i}^{d} w_i w_j} \\
&= \frac{v + v \rho^{x}(d(\cos\theta)^2 - 1)}{\frac{1}{p}(c^2+v) - c^2 + v \rho^{x}(d(\cos\theta)^2 - 1)},
\end{aligned} \tag{4}$$