Tensor Switching Networks: Supplementary Material

Chuan-Yung Tsai[*], Andrew Saxe[*], David Cox
Center for Brain Science, Harvard University, Cambridge, MA 02138
{chuanyungtsai,asaxe,davidcox}@fas.harvard.edu

Alternative Derivation of TS-ReLU Kernel

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_0}$, we wish to derive $\bar{k}^{\text{TS}}$ in

$$k_1^{\text{TS}}(\mathbf{x}, \mathbf{y}) = 2\,\mathbb{E}\left[H\left(\mathbf{w}^\intercal\mathbf{x}\right) H\left(\mathbf{w}^\intercal\mathbf{y}\right)\right]\mathbf{x}^\intercal\mathbf{y} = \underbrace{2\,P\left(\mathbf{w}^\intercal\mathbf{x} > 0 \text{ and } \mathbf{w}^\intercal\mathbf{y} > 0\right)}_{\bar{k}^{\text{TS}}}\mathbf{x}^\intercal\mathbf{y},$$

where $\mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$. To achieve this goal, we define

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{x}^\intercal/\sigma\left\|\mathbf{x}\right\| \\ \mathbf{y}^\intercal/\sigma\left\|\mathbf{y}\right\| \end{bmatrix}}_{\mathbf{L}} \mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \underbrace{\begin{bmatrix} 1 & \cos\theta \\ \cos\theta & 1 \end{bmatrix}}_{\mathbf{L}(\sigma^2\mathbf{I})\mathbf{L}^\intercal}\right). \tag{1}$$

Then we have

$$\begin{aligned}
\bar{k}^{\text{TS}} &= 2\,P\left(\mathbf{w}^\intercal\mathbf{x} > 0 \text{ and } \mathbf{w}^\intercal\mathbf{y} > 0\right) \\
&= 2\,P\left(\frac{\mathbf{w}^\intercal\mathbf{x}}{\sigma\left\|\mathbf{x}\right\|} > 0 \text{ and } \frac{\mathbf{w}^\intercal\mathbf{y}}{\sigma\left\|\mathbf{y}\right\|} > 0\right) \\
&= 2\,P\left(z_1 > 0 \text{ and } z_2 > 0\right) \\
&= 2\int_0^\infty\int_0^\infty \frac{1}{2\pi\sqrt{1-\cos^2\theta}} \exp\left(-\frac{z_1^2 - 2z_1z_2\cos\theta + z_2^2}{2\left(1-\cos^2\theta\right)}\right) dz_1 dz_2 && \text{Using PDF of (1)} \\
&= \frac{1}{\pi\sin\theta}\int_0^{\frac{\pi}{2}}\int_0^\infty r\exp\left(-r^2\frac{1-\cos\theta\sin 2\phi}{2\sin^2\theta}\right) dr d\phi && \text{Polar Coordinates} \\
&= \frac{1}{\pi\sin\theta}\int_0^{\frac{\pi}{2}}\left(-\frac{1}{2a}\exp\left(-r^2a\right)\Big|_{r=0}^{\infty}\right) d\phi && a = \frac{1-\cos\theta\sin 2\phi}{2\sin^2\theta} \\
&= \frac{1}{\pi\sin\theta}\int_0^{\frac{\pi}{2}}\frac{1}{2a} d\phi \\
&= \frac{1}{\pi}\sin\theta\int_0^{\frac{\pi}{2}}\frac{1}{1-\cos\theta\sin 2\phi} d\phi && \text{Special Case of (A.3) of [1]} \\
&= \frac{1}{\pi}\sin\theta\left(\frac{\pi-\theta}{\sin\theta}\right) && \text{Following (A.6) of [1]} \\
&= 1 - \frac{\theta}{\pi}.
\end{aligned}$$

Thus, $k_1^{\text{TS}}(\mathbf{x}, \mathbf{y}) = \left(1 - \frac{\theta}{\pi}\right)\mathbf{x}^\intercal\mathbf{y}$.

[*]Equal contribution.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

References

[1] Y. Cho and L. Saul, “Large-Margin Classification in Infinite Neural Networks,” *Neural Computation*, 2010.