

---

# Dropout Distillation Supplementary Material

---

**Samuel Rota Bulò**  
FBK-irst, Trento, Italy

ROTABULO@FBK.EU

**Lorenzo Porzi**  
FBK-irst, Trento, Italy

PORZI@FBK.EU

**Peter Kotschieder**  
Mapillary, Graz, Austria - Microsoft Research, Cambridge, UK

PKONTSCHIEDER@MAPILLARY.COM

## Abstract

This document provides the proof of Theorem 1 in the main paper (Rota Bulò et al., 2016).

*Proof of Thm.1.* The proof of *i)* follows by simple application of Jensen’s inequality given the convexity assumption of  $\ell$ :

$$\begin{aligned} J(q) &= \mathbb{E}_{\mathbf{x}} [\ell(f_{\text{dropout}}(\mathbf{x}), q(\mathbf{x}))] \\ &\leq \mathbb{E}_{\mathbf{x}, \sigma} [\ell(f_{\Theta^*, \sigma}(\mathbf{x}), q(\mathbf{x}))] = J'(q). \end{aligned}$$

As for *ii)*, by exploiting the expression for  $\ell$  in terms of  $g_1, g_2, g_3$  we have

$$\begin{aligned} J(q) &= \mathbb{E}_{\mathbf{x}} [g_1(f_{\text{dropout}}(\mathbf{x})) + g_2(q(\mathbf{x})) \\ &\quad + f_{\text{dropout}}(\mathbf{x})^\top g_3(q(\mathbf{x}))] \\ &= \mathbb{E}_{\mathbf{x}, \sigma} [g_1(f_{\Theta^*, \sigma}(\mathbf{x})) + g_2(q(\mathbf{x})) \\ &\quad + f_{\Theta^*, \sigma}(\mathbf{x})^\top g_3(q(\mathbf{x}))] \\ &\quad - \underbrace{\mathbb{E}_{\mathbf{x}, \sigma} [g_1(f_{\Theta^*, \sigma}(\mathbf{x}))]}_{\Delta} + \mathbb{E}_{\mathbf{x}} [g_1(f_{\text{dropout}}(\mathbf{x}))] \\ &= \mathbb{E}_{\mathbf{x}, \sigma} [\ell(f_{\Theta^*, \sigma}(\mathbf{x}), q(\mathbf{x}))] + \Delta \\ &= J'(q) + \Delta, \end{aligned}$$

where  $\Delta$  is a constant that does not depend on  $q$ .

The same derivations hold true if we replace  $\mathbb{E}_{\mathbf{x}} [\cdot]$  with an empirical average.  $\square$

## References

Rota Bulò, S., Porzi, L., and Kotschieder, P. Dropout distillation. In *Int. Conf. on Machine Learning*, 2016.

*Proceedings of the 33<sup>rd</sup> International Conference on Machine Learning*, New York, NY, USA, 2016. JMLR: W&CP volume 48. Copyright 2016 by the author(s).