

---

# Gated Softmax Classification – Errata

---

**Roland Memisevic**  
Department of Computer Science  
ETH Zurich  
Switzerland  
roland.memisevic@gmail.com

## 1 Errata

Eq. (11) in [1] is incorrect, and should be written correctly

$$\frac{\partial \log p(y|\mathbf{x})}{\partial W_{\bar{y}ik}} = (\delta_{\bar{y}y} - p(\bar{y}|\mathbf{x})) \sigma\left(\sum_i x_i W_{\bar{y}ik}\right) x_i \quad \text{with} \quad \sigma(a) = (1 + \exp(-a))^{-1} \quad (1)$$

The gradients of the factored version of the model stay the same:

$$\frac{\partial \log p(y|\mathbf{x})}{\partial W_{if}^{\mathbf{x}}} = \sum_{\bar{y},k} \frac{\partial \log p(y|\mathbf{x})}{\partial W_{\bar{y}ik}} \frac{\partial W_{\bar{y}ik}}{\partial W_{if}^{\mathbf{x}}} \quad (2)$$

Similarly for  $W_{yf}^{\mathbf{y}}$  and  $W_{kf}^{\mathbf{h}}$  (with the sums running over the remaining indices).

## References

- [1] R. Memisevic, C. Zach, G. Hinton, and M. Pollefeys. Gated softmax classification. In *Advances in Neural Information Processing Systems 23*. MIT Press, 2010.

## Acknowledgments

We thank Laurens van der Maaten for pointing out the mistake.