Errata: Practical Uncertainty in Neural Networks

Ethan Goan

p37, Eqn 3.30 for leap frog integration, the step should be $\epsilon/2$. I am not happy with the notation there. Instead use,

$$\begin{aligned} v_i(t+\frac{\epsilon}{2}) &= v_i(t) + \frac{\epsilon}{2}\frac{dv_i(t)}{dt}, \\ w_i(t+\epsilon) &= w_i(t) + \epsilon\frac{dw_i(t+\epsilon/2)}{dt}, \\ v_i(t+\epsilon) &= v_i(t+\frac{\epsilon}{2}) + \frac{\epsilon}{2}\frac{dv_i(t+\epsilon)}{dt}. \end{aligned}$$

p51 For the α divergence on page 75, the limits to relate to forward/reverse KL divergence and Hellinger distance were not correct. It should be, $KL(q|p) : \alpha \to 0, KL(p|q) : \alpha \to 1$, Hellinger: $\alpha \to 0.5$.

p51 Eq. 3.58 missing factor of 1/2

p107, Eq 4.7: $p(\mathbf{w}|\mathbf{X},\mathbf{Y},\theta) \propto \Big[\prod_{i=0}^{N-1}\mathcal{N}\Big(\mathbf{y}|f(\mathbf{x}_i;\theta)\mathbf{w},\sigma^2\Big)\Big]p(\mathbf{w})$.

Eq 4.11 and 4.15: are missing Φ, they should read
$\sigma_N^2(\hat{\mathbf{x}}) = \sigma^2 + \Phi(\hat{\mathbf{x}})^T\Sigma_\pi\Phi(\hat{\mathbf{x}})$ and $\sigma_N^2(\hat{\mathbf{x}}) \approx \sigma^2\Phi(\hat{\mathbf{x}})^T\Sigma_{\text{diag}}\Phi(\hat{\mathbf{x}})$

p111, Eq 4.22: The approximation from ref [34] in the paper provides the Taylor Series approximation of $\mathcal{N}\Big(\frac{\mu_j}{\mu_d}, \frac{\mu_j^2}{\mu_d^2}\big(\sigma_j^2/\mu_j^2 + \sigma_d^2/\mu_d^2\big)\Big)$. The issue with this is the square of the μ_d terms, as this is at risk of numerical overflow in half floating-point precision. Instead we use in $\sigma^2 \approx \mu_j/\mu_d\sqrt{\sigma_j^2+\sigma_d^2}$, which emperically follwed a similar trend and was less of a risk for numerical overflow (though not immune).

p135, Eq. 5.13, the integral sign should be there in the denominator.

p143, Eqn 5.43, and **Pg. 146** Eqn 5.46, Typo in missing square term.

$$\begin{aligned} &\frac{\sigma^2}{2}\int_{-\infty}^{\infty} v^2 \operatorname{erf}\Big(\frac{(\sigma v+\mu)}{\sqrt{2}}\Big)\phi(v)dv \\ =&\frac{\sigma^2}{4\sqrt{2\pi}}\Big[\frac{\sqrt{\pi}}{\alpha_1\sqrt{\alpha_1}}\operatorname{erf}\Big(\frac{\sqrt{\alpha_1}b_1}{\sqrt{\alpha_1+a_1^2}}\Big) - \frac{1}{\alpha_1\sqrt{\alpha_1+a_1^2}}\frac{2a_1^2b_1}{\alpha_1+a_1^2}\exp\big(\frac{-\alpha_1 b_1^2}{\alpha_1+a_1^2}\big)\Big] \end{aligned}$$

This same typo carried over to integral property from [240, EQ. 2.7.3.9]

$$\begin{aligned} \int_{-\infty}^{\infty} z^2\exp(\alpha^2)\operatorname{erf}(\alpha_1 z+\beta_1) = &\frac{\sqrt{\pi}}{2\alpha\sqrt{\alpha}}\operatorname{erf}\Big(\frac{\sqrt{\alpha}\beta_1}{\sqrt{\alpha+\alpha_1^2}}\Big) \\ &- \frac{1}{\sqrt{\alpha+\alpha_1^2}}\frac{2\alpha_1^2\beta_1}{\alpha+\alpha_1^2}\exp\frac{-\alpha\beta_1^2}{\alpha+\alpha^2} \end{aligned}$$

And to the final result on **Pg. 147**, Eqn 5.61 for the approximation to the variance of the Gelu function,

$$
\begin{aligned}
\sigma_g^2 \approx & \frac{\sigma^2+\mu^2}{2} \\
& + \frac{\sigma^2}{4\sqrt{2\pi}}\Big[\frac{\sqrt{\pi}}{\alpha_1\sqrt{\alpha_1}}\,\mathrm{erf}\Big(\frac{\sqrt{\alpha_1}b_1}{\sqrt{\alpha_1+a_1^2}}\Big) - \frac{1}{\alpha_1\sqrt{\alpha_1+a_1^2}}\frac{2a_1^2b_1}{\alpha_1+a_1^2}\exp\big(\frac{-\alpha_1 b_1^2}{\alpha_1+a_1^2}\big)\Big] \\
& + \frac{\mu\sigma a_2}{\alpha_2\sqrt{2\pi(a_2^2+\alpha_2)}}\exp\Big(-\frac{\alpha_2 b_2^2}{a_2^2+\alpha_2}\Big) \\
& + \frac{\mu^2}{2\sqrt{2\alpha_3}}\,\mathrm{erf}\Big(\frac{2\alpha_3 b_3}{2\sqrt{\alpha_3^2+\alpha_3 a_3^2}}\Big) \\
& - \frac{(2\alpha_4^2+b_4^2)}{16\sigma\alpha_4^5\sqrt{2}}\exp\Big(\frac{b_4^2}{4\alpha_4^2}+\gamma\Big) - \mu_g^2
\end{aligned}
$$

where, $\alpha_1, \alpha_2, \alpha_3 = \frac{1}{2}$, $b_1, b_2, b_3 = \frac{\mu}{\sqrt{2}}$, $a_1, a_2, a_3 = \frac{\sigma}{\sqrt{2}}$, $\alpha_4 = \sqrt{\frac{1}{2\sigma^2} + \frac{\eta}{2}}$, $b_4 = \frac{\mu}{\sigma^2}$ and $\gamma = -\frac{\mu^2}{2\sigma^2}$, $\eta = 1.23907$.

pg. 150, Eqn 5.68, the last term are wrong way around. Should read,

$$\Omega = \Omega_X\Omega_Y + \Omega_X\mu_Y^2 + \mu_X^2\Omega_Y.$$