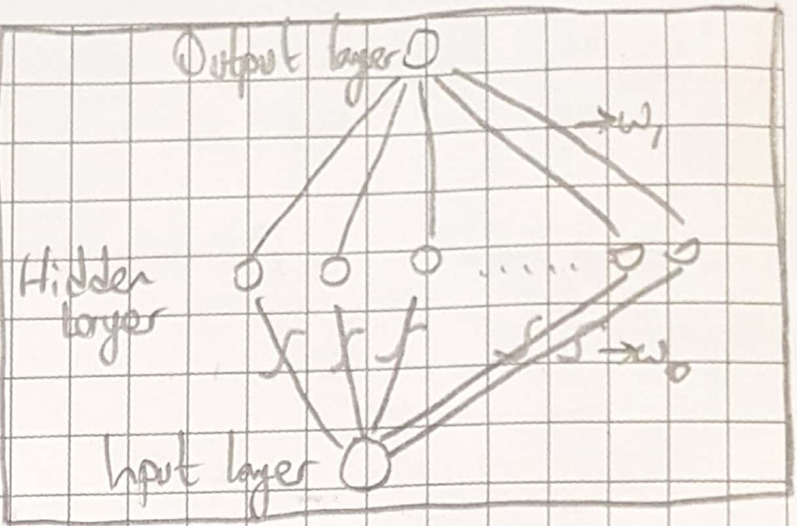
1- $\varepsilon_w \sim N(0,1)$ for weights

2- $\sigma_w = \ln(1+e^{\rho_w})$ for weights



3- $w = \mu_w + \ln(1+e^{\rho_w}) \otimes \varepsilon_w$

4- $$\frac{\partial w}{\partial \mu_w} = 1$$

5- $$\frac{\partial w}{\partial \rho_w} = \varepsilon_w \otimes \frac{1}{1+e^{\rho_w}} \otimes e^{\rho_w}$$

6- $$\mathrm{sigmoid}(k) = \frac{1}{1+e^{-k}}$$

7- $$\frac{d\,\mathrm{sigmoid}(k)}{\partial k} = (-1) \cdot \frac{1}{(1+e^{-k})^2} \cdot e^{-k} = \frac{1}{1+e^{-k}} \cdot \frac{-e^{-k}}{1+e^{-k}}$$

$$= \mathrm{sigmoid}(k) \cdot (1-\mathrm{sigmoid}(k))$$

8- Say $m \sim N(\mu, \sigma^2)$, then

$$f(m) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{\frac{-(m-\mu)^2}{2\sigma^2}} \rightarrow \text{pdf of normal dist.}$$

$$\ln(f(m)) = -\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{(m-\mu)^2}{2\sigma^2}$$

9- $$\mathrm{Loss} = F(w,\theta) = \underbrace{\ln(q(w|\theta))}_{\text{Logarithm of variational posterior probability}} - \underbrace{\ln(P(w))}_{\text{Logarithm of prior probability}} - \underbrace{\ln(P(D|w))}_{\text{Logarithm of likelihood}}, \text{ where } \theta = (\mu, \rho)$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y-\hat{y})^2$$ where N is the sample size

$$q(w|\theta) \sim N(\mu_w, \sigma_w^2)$$

$$\ln(q(w|\theta)) = -\ln\sqrt{2\pi} - \ln\sigma_w - \frac{(w-\mu_w)^2}{2\sigma_w^2}$$

$$= -\ln\sqrt{2\pi} - \ln(\ln(1+e^{\rho_w})) - \frac{(\cancel{\mu_w} + \cancel{\ln(1+e^{\rho_w})} \otimes \varepsilon_w - \cancel{\mu_w})^2}{2 \cdot \cancel{(\ln(1+e^{\rho_w}))^2}}$$

$$= -\ln\sqrt{2\pi} - \ln(\ln(1+e^{\rho_w})) - \frac{\varepsilon_w \otimes \varepsilon_w}{2}$$

$$P(w) \sim N(0,1)$$

$$\ln(P(w)) = -\ln\sqrt{2\pi} - \cancel{\ln 1} - \frac{(w-0)^2}{2 \cdot 1^2}$$

$$= -\ln\sqrt{2\pi} - \frac{w^2}{2}$$

$$= -\ln\sqrt{2\pi} - \frac{(\mu_w + \ln(1+e^{\rho_w}) \otimes \varepsilon_w)^2}{2}$$

$$P(D|w) \sim N(\hat{y}, tol) \quad \text{where} \quad \hat{y} = \text{sigmoid}(x . w_0^T) . w_1^T$$

$$\ln(P(D|w)) = -\ln\sqrt{2\pi} - \ln(tol) - \frac{(y-\hat{y})^2}{2 tol^2}$$

$$= -\ln\sqrt{2\pi} - \ln(tol) - \frac{(y - \text{sigmoid}(x . w_0^T) . w_1^T)^2}{2 tol^2}$$

$$= -\ln\sqrt{2\pi} - \ln(tol) - \frac{\left[y - \text{sigmoid}(x . (\mu_{w_0} + \ln(1+e^{\rho_{w_0}}) \otimes \varepsilon_{w_0})^T) . (\mu_{w_1} + \ln(1+e^{\rho_{w_1}}) \otimes \varepsilon_{w_1})^T\right]^2}{2 tol^2}$$

10- $$\frac{\partial \ln(P(D|\omega))}{\partial \hat{y}} = \frac{+\cancel{2}(y-\hat{y})}{\cancel{2}\, tol^2}\cdot(+1) = \frac{y-\hat{y}}{tol^2}$$

11- $$\frac{\partial \hat{y}}{\partial \omega_1} = \text{sigmoid}(x.\omega_0^T)$$

12- $$\frac{\partial Loss}{\partial \mu_{\omega_1}} = \frac{\partial log\_var\_post}{\partial \mu_{\omega_1}} - \frac{\partial log\_prior}{\partial \mu_{\omega_1}} - \frac{\partial log\_likelihood}{\partial \mu_{\omega_1}}$$

$$= \frac{\partial \ln(q(\omega_1|\theta))}{\partial \mu_{\omega_1}} - \frac{\partial \ln(P(\omega_1))}{\partial \mu_{\omega_1}} - \frac{\partial \ln(P(D|\omega_1))}{\partial \mu_{\omega_1}}$$

$$\frac{\partial \ln(q(\omega_1|\theta))}{\partial \mu_{\omega_1}} = 0$$

$$\frac{\partial \ln(P(\omega_1))}{\partial \mu_{\omega_1}} = -\frac{\cancel{2}(\mu_{\omega_1} + \ln(1+e^{l\omega_1}) \otimes \varepsilon_{\omega_1})}{\cancel{2}} \otimes 1 = -\omega_1$$

$$\frac{\partial \ln(P(D|\omega_1))}{\partial \mu_{\omega_1}} = \frac{\partial \ln(P(D|\omega))}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \omega_1} \cdot \frac{\partial \omega_1}{\partial \mu_{\omega_1}}$$

$$= \frac{y-\hat{y}}{tol^2} \cdot \text{sigmoid}(x.\omega_0^T) \otimes 1$$

13- $$\frac{\partial Loss}{\partial \mu_{\omega_0}} = \frac{\partial \ln(q(\omega_0|\theta))}{\partial \mu_{\omega_0}} - \frac{\partial \ln(P(\omega_0))}{\partial \mu_{\omega_0}} - \frac{\partial \ln(P(D|\omega_0))}{\partial \mu_{\omega_0}}$$

$$\frac{\partial \ln(q(\omega_0|\theta))}{\partial \mu_{\omega_0}} = 0$$

$$\frac{\partial \ln(P(\omega_0))}{\partial \mu_{\omega_0}} = \frac{-\cancel{2}(\mu_{\omega_0} + \ln(1+e^{l\omega_0}) \otimes \varepsilon_{\omega_0})}{\cancel{2}} \otimes 1 = -\omega_0$$

$$\frac{\partial \ln(P(D|w_0))}{\partial \mu_{w_0}} = \frac{\partial \ln(P(D|w_0))}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h} \cdot \frac{\partial h}{\partial z} \cdot \frac{\partial z}{\partial w_0} \cdot \frac{\partial w_0}{\partial \mu_{w_0}}$$

where $\hat{y} = \underbrace{\text{sigmoid}(\underbrace{x.w_0^T}_{z})}_{h} . w_1^T$

$$= \frac{y-\hat{y}}{\sigma_0^2} . w_1 \otimes \left[\text{sigmoid}(x.w_0^T).(1-\text{sigmoid}(x.w_0^T))\right].X \otimes 1$$

4 $$\frac{\partial Loss}{\partial \rho_{w_1}} = \frac{\partial \ln(q(w_1|\theta))}{\partial \rho_{w_1}} - \frac{\partial \ln(P(w_1))}{\partial \rho_{w_1}} - \frac{\partial \ln(P(D|w_1))}{\partial \rho_{w_1}}$$

$$\frac{\partial \ln(q(w_1|\theta))}{\partial \rho_{w_1}} = -\frac{1}{\ln(1+e^{\rho_{w_1}})} \otimes \frac{1}{1+e^{\rho_{w_1}}} \otimes e^{\rho_{w_1}}$$

$$\frac{\partial \ln(P(w_1))}{\partial \rho_{w_1}} = \frac{-\cancel{2}(\mu_{w_1} + \ln(1+e^{\rho_{w_1}}) \otimes \varepsilon_{w_1})}{\cancel{2}} \otimes \frac{\partial w_1}{\partial \rho_1}$$

$$= -w_1 \otimes \left[\frac{1}{1+e^{\rho_{w_1}}} \otimes e^{\rho_{w_1}} \otimes \varepsilon_{w_1}\right]$$

$$\frac{\partial \ln(P(D|w_1))}{\partial \rho_{w_1}} = \frac{\partial \ln(P(D|w_1))}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \cdot \frac{\partial w_1}{\partial \rho_{w_1}}$$

$$= \left(\frac{y-\hat{y}}{\sigma_0^2}\right) . \text{sigmoid}(x.w_0^T) \otimes \left[\frac{1}{1+e^{\rho_{w_1}}} \otimes e^{\rho_{w_1}} \otimes \varepsilon_{w_1}\right]$$

15- $$\frac{\partial Loss}{\partial \rho_{w_0}} = \frac{\partial \ln(q(w_0|\theta))}{\partial \rho_{w_0}} - \frac{\partial \ln(P(w_0))}{\partial \rho_{w_0}} - \frac{\partial \ln(P(D|w_0))}{\partial \rho_{w_0}}$$

$$\frac{\partial \ln(q(w_0|\theta))}{\partial \rho_{w_0}} = -\frac{1}{\ln(1+e^{\rho_{w_0}})} \otimes \frac{1}{1+e^{\rho_{w_0}}} \otimes e^{\rho_{w_0}}$$

$$\frac{\partial \ln(P(w_0))}{\partial \rho_{w_0}} = \frac{-\cancel{2}(\mu_{w_0} + \ln(1+e^{\rho_{w_0}}) \otimes \varepsilon_{w_0})}{\cancel{2}} \otimes \frac{\partial w_0}{\partial \rho_{w_0}}$$

$$= -w_0 \otimes \left[\frac{1}{1+e^{\rho_{w_0}}} \otimes e^{\rho_{w_0}} \otimes \varepsilon_{w_0}\right]$$

$$\frac{\partial \ln(P(D|w_0))}{\partial \rho_{w_0}} = \frac{\partial \ln(P(D|w_0))}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h} \cdot \frac{\partial h}{\partial z} \cdot \frac{\partial z}{\partial w_0} \cdot \frac{\partial w_0}{\partial \rho_{w_0}}$$

where $$\hat{y} = \underbrace{\text{sigmoid}(\underbrace{x.w_0^T}_{z})}_{h} . w_1^T$$

$$= \frac{y-\hat{y}}{\sigma^2} . w_1 \otimes \left[\text{sigmoid}(x.w_0^T) \otimes (1-\text{sigmoid}(x.w_0^T))\right] . x \otimes \left[\frac{e^{\rho_{w_0}}}{1+e^{\rho_{w_0}}} \otimes \varepsilon_{w_0}\right]$$

16
$$\mu_{w_1} = \mu_{w_1} - \alpha . \frac{\partial Loss}{\partial \mu_{w_1}}$$

$$\mu_{w_0} = \mu_0 - \alpha . \frac{\partial Loss}{\partial \mu_{w_0}}$$

$$\rho_{w_1} = \rho_{w_1} - \alpha . \frac{\partial Loss}{\partial \rho_{w_1}}$$

$$\rho_{w_0} = \rho_{w_0} - \alpha . \frac{\partial Loss}{\partial \rho_{w_0}}$$

where α is the learning rate

17- $\hat{y} = \text{sigmoid}(x.w_0^T).w_1^T$

18- $y = 10.\sin(2\pi x) + \varepsilon$