Effects of change of units on training procedure

- energy $= [A]$ positions $= [B]$ forces $= [A/B]$
- Assume change of units (☆) as follows:

$$1A = a\cdot\tilde{A}\,,\ 1B = b\cdot\tilde{B} \Rightarrow 1\frac{A}{B} = \frac{a}{b}\frac{\tilde{A}}{\tilde{B}} \quad \text{with } a,b\in R$$

- Datapoint $d_i = (p_i, e_i, f_i)$ with $p_i, f_i \in R^{N_{at}\times 3}$, $e_i \in R$
- Training on the original units, entails calculation of

$$MSE = \frac{\lambda_E}{n}\sum_{i=1}^{n}\left(e_i - \hat{E}(p_i)\right)^2 + \frac{\lambda_F}{n}\sum_{i=1}^{n}\left(f_i - \hat{F}(p_i)\right)^2 \quad \text{(unitless)}$$

trained on org. units

- Training on other units (☆) leads to

$$\widetilde{MSE} = \frac{\lambda_E}{n}\sum_{i=1}^{n}\left(a\, e_i - \hat{E}'(b\cdot p_i)\right)^2 + \frac{\lambda_F}{n}\sum_{i=1}^{n}\left(\frac{a}{b} f_i - \hat{F}'(b\cdot p_i)\right)^2$$

- Let's assume $\hat{E}'(b\cdot p_i) = a\,\hat{E}(p_i)$ & $\hat{F}'(b\cdot p_i) = \frac{a}{b}\hat{F}(p_i)$ (⌂)

$$\Rightarrow \widetilde{MSE} = a^2\frac{\lambda_E}{n}\sum_{i=1}^{n}\left(e_i - \hat{E}(p_i)\right)^2 + \left(\frac{a}{b}\right)^2\frac{\lambda_F}{n}\sum_{i=1}^{n}\left(f_i - \hat{F}(p_i)\right)^2$$

$$= MSE(\tilde{\lambda}_1, \tilde{\lambda}_2) \quad \& \quad \tilde{\lambda}_1 = a^2\lambda_1\,,\ \tilde{\lambda}_2 = \frac{a^2}{b^2}\lambda_2$$

$$\Rightarrow \frac{\tilde{\lambda}_E}{\tilde{\lambda}_F} = \frac{a^2\lambda_E}{\frac{a^2}{b^2}\lambda_F} = b^2\frac{\lambda_E}{\lambda_F} \neq \frac{\lambda_E}{\lambda_F} \quad \& \quad \tilde{\lambda}_E + \tilde{\lambda}_F \neq 1$$

$\Rightarrow$ changing units leads to different weighting of the training targets

$\Rightarrow$ Changing the hyperparameter associated with free choice when changing units is equivalent to adjusting weights λ_1 & λ_2

- Is the assumption (⌂) valid?
- Theoretically, model is able to learn such "simple scaling" of the data (for two completely separate models the assumption should be valid also in practise)
- The change in weights resulting from the assumption leads to the fact, that the assumption doesn't hold in practise
- Maybe the result can be used as a thumb rule as illustrated below
- Note, that the above argument changes for different losses

Test case: From Hartree to kcal/mol

- data given in Angstrom, Hartree, fs & Angstrom, kcal/mol, fs
- original weights $\lambda_E = 0.01$ & $\lambda_F = 0.99$ for kcal/mol
- $1 \frac{\text{kcal}}{\text{mol}} \overset{\text{Wiki}}{=} 4.184 \frac{\text{kJ}}{\text{mol}} \overset{\text{Wiki}}{=} 4.184 \frac{10^3 \text{ J}}{6.022\,140\,76 \cdot 10^{23}}$

$$= 6.947\,695\,457\,06 \cdot 10^{-21} \text{ J}$$

$$1 \text{ Hartree} \overset{\text{Wiki}}{=} 4.359\,744\,722\,2060(48) \times 10^{-18} \text{ J}$$

$$= 627.509\,474\,062\,48 \frac{\text{kcal}}{\text{mol}} \equiv a^{-1} \cdot \frac{\text{kcal}}{\text{mol}}$$

$$\Rightarrow 1 \frac{\text{kcal}}{\text{mol}} = 0.001\,593\,601\,44 \text{ Hartree} \equiv a \cdot \text{Hartree}$$

- Choose λ'_F, λ'_E such that

$$\text{i)} \quad \frac{\tilde{\lambda}_F}{\tilde{\lambda}_E} = b^{-2} \frac{\lambda'_F}{\lambda'_E} = \frac{\lambda_F}{\lambda_E}$$

$$\text{ii)} \quad \tilde{\lambda}_F + \tilde{\lambda}_E = \frac{a^2}{b^2} \lambda'_F + a^2 \lambda'_E = 1$$

or c, if $\lambda_F + \lambda_E = c$

where i) is given, since we only change unit of energy, i.e., $b = 1$

- Notation:
  - λ: original weights
  - $\tilde{\lambda}$: effective weights resulting from change of units
  - λ': weights that have to be set, such that effective weights $\tilde{\lambda}$ will have same proportion as original weights λ & sum to one (or same const. c)

$\cdot$ ad ii) $\tilde{\lambda}_F = a^2 \lambda'_F$ & $\tilde{\lambda}_E = a^2 \lambda'_E$

$$\Rightarrow \text{ set } \lambda'_F = \frac{\lambda_F}{a^2} \quad \& \quad \lambda'_E = \frac{\lambda_E}{a^2}$$

$$\Rightarrow \tilde{\lambda}_F + \tilde{\lambda}_E = a^2 \lambda'_F + a^2 \lambda'_E = \lambda_F + \lambda_E \; (= 1)$$

$$\lambda'_E = 0.01 \cdot a^{-2} \cong 0.01 \cdot (627.509\,474\,062\,48)^2$$

$$= 3\,937.681\,400\,381\,7 \cong 3\,937.68$$

$$\lambda'_F = 0.99 \cdot a^{-2} \cong 389\,830.458\,637\,788\,6$$

$$\cong 389\,830.46$$

Test case: From atomic units to Ang & kcal/mol

- 1 Hartree $\cong$ 627.509 $\frac{kcal}{mol}$ & 1 Bohr $\cong$ 0.529 177 Å = b · Å
- Schnetpack default $\lambda_E = 0.01$ $\lambda_F = 0.99$

$$\rightarrow \left(\frac{\lambda_E}{\lambda_F}\right)^{-1} = 9.9$$ (more compact result with inverse)

$$\xrightarrow[(\star)]{\text{conv}} \left(\frac{\tilde{\lambda}_E}{\tilde{\lambda}_F}\right)^{-1} = \left(0.529\,177^2 \frac{\lambda_E}{\lambda_F}\right)^{-1} = 3\,535.356\,995 \neq 9.9 \ \lightning$$

- choose new λ'_E & λ'_F such that

$$\left(\frac{\tilde{\lambda}_E}{\tilde{\lambda}_F}\right)^{-1} = b^{-2} \frac{\lambda'_F}{\lambda'_E} \overset{!}{=} 9.9 \quad (\star) \quad \text{with}$$

$$\lambda'_E + \lambda'_F = 1 \iff \lambda'_E = 1 - \lambda'_F$$

$$\overset{(\star)}{\Rightarrow} \lambda'_F = 9.9\, b^2 \cdot \lambda'_E = 9.9\, b^2 \cdot (1 - \lambda'_F)$$

$$\iff \lambda'_F \left(1 + 9.9\, b^2\right) = 9.9\, b^2$$

$$\iff \lambda'_F = \frac{9.9\, b^2}{1 + 9.9\, b^2} = 0.734\,908\,341\,39$$

$$\Rightarrow \lambda'_E = 0.265\,091\,658\,61$$

- Thus, choose $\lambda'_F \cong 0.73$ & $\lambda'_E \cong 0.27$ when training on kcal/mol & Å

Transformations & Physics

- In a closed system of N atom with only conservative forces present, we have the following relationship between the potential energy E and the force F:

$$\vec{F} = -\vec{\nabla}_{\vec{r}} E(\vec{r}) \quad \text{or} \quad (\vec{F})_\alpha = -\frac{\partial}{\partial r_\alpha} E(\vec{r}) \qquad (\mathcal{F})$$

with $\vec{F} \in \mathbb{R}^3$, $E \in \mathbb{R}$ & $\alpha = 1,2,3 \hat{=} x,y,z$

- Let $\vec{a}_j$ be the position of the j-th atom

$$\Rightarrow \vec{F}(\vec{a}_j) = -\vec{\nabla}_{\vec{r}} E(\vec{r})\Big|_{\vec{r}=\vec{a}_j} \quad \text{or} \quad (\vec{F}(\vec{a}_j))_\alpha = -\frac{\partial}{\partial r_\alpha} E(\vec{r})\Big|_{\vec{r}=\vec{a}_j}$$

- Under which transformations on $E, \vec{r}$ & $\vec{F}$ is $(\mathcal{F})$ unchanged, i.e.,

$$-\frac{\partial}{\partial r_\alpha} t_E(E(t_r(\vec{r}))) = t_F(\vec{F}_\alpha)$$

- Examine addition & multiplication of constants, due to their importance for standardization & normalization

Transformations of the energy

- $t(E) = E + a, \quad a \in \mathbb{R}$

$$\Rightarrow \frac{\partial}{\partial r_\alpha}(t(E)) = \frac{\partial}{\partial r_\alpha} E(\vec{r}) + \cancelto{0}{\frac{\partial}{\partial r_\alpha} a} = \frac{\partial}{\partial r_\alpha} E(\vec{r})$$

$\Rightarrow (\mathcal{F})$ inv. under addition of const. to energy

- $t(E) = b \cdot E \quad , \quad b \in \mathbb{R}$

$$\Rightarrow \frac{\partial}{\partial r_\alpha} t(E) = b \frac{\partial}{\partial r_\alpha} E(\vec{r})$$

$\Rightarrow$ (F) <u>not</u> inv. under mult. of const. with energy

$\Rightarrow$ (F) inv. if both E & $\vec{F}$ are multiplied by const.

- One cannot standardize or normalize both E & F without impacting physics, i.e. (F)

Transformations of the force

- $t_F(F) = F + \vec{a}$ with $\vec{a} \in \mathbb{R}^3$

$$\Rightarrow \frac{-\partial}{\partial r_\alpha} E = F_\alpha \neq F_\alpha + a_\alpha$$

$\Rightarrow$ only invariant for $\tilde{E}(\vec{r}) = E + \vec{r} \cdot \vec{a}$

- $t_F(F) = b \cdot F$ is analog to energy

Transformations of the position

- $t_r(\vec{r}) = \vec{r} + \vec{a} \quad , \quad \vec{a} \in \mathbb{R}^3$

$$\Rightarrow -\frac{\partial}{\partial r_\alpha} E(\vec{r} + \vec{a}) = -\frac{\partial}{\partial r'_\alpha} E(\vec{r}') = F_\alpha(\vec{r}')$$

subst. $r'_\alpha = r_\alpha + a_\alpha \rightarrow \frac{dr'_\alpha}{dr_\alpha} = 1$

$\Rightarrow$ subtraction of center of mass leaves the physics unchanged

- $t_r(\vec{r}) = \vec{r} \cdot b \quad , \; b \in \mathbb{R}$

$$\Rightarrow -\frac{\partial}{\partial r_\alpha} E(\vec{r} \cdot b) = -\frac{1}{b} \frac{\partial}{\partial r'_\alpha} E(\vec{r}') = \frac{1}{b} F_\alpha(\vec{r}')$$

subst. $r'_\alpha = r_\alpha \cdot b \Rightarrow \frac{dr'_\alpha}{dr_\alpha} = b$

$\Rightarrow$ scaling positions leads to scaling of forces, which is exactly what's happening when changing units

Conclusions

- Centering energy & positions leaves everything unchanged (SchNetPack default)
- When standardizing energies, one has to correct the forces accordingly

Use case: Standardize energy & positions

- giv.: $\mu_E, \vec{\mu}_p, \sigma_E, \vec{\sigma}_p$ all computed on the training data

$$\Rightarrow t_E(E) = \frac{E - \mu_E}{\sigma_E} \equiv E' \quad \& \quad t_p(\vec{r}) = \frac{\vec{r} - \vec{\mu}_p}{\vec{\sigma}_p} \equiv \vec{r}'$$

- Since addition of const. leaves (F) unchanged we write

$$E' = \frac{\tilde{E}}{\sigma_E} \quad \& \quad \vec{r}' = \frac{\tilde{\vec{r}}}{\vec{\sigma}_p}$$

$$\Rightarrow -\frac{\partial}{\partial r_\alpha} E'(\vec{r}') = -\frac{\partial}{\partial r_\alpha}\left(\frac{\tilde{E}(\frac{\tilde{\vec{r}}}{\vec{\sigma}_p})}{\sigma_E}\right) = -\left(\frac{\vec{\sigma}_p}{\sigma_E}\right) \frac{\partial}{\partial r'_\alpha} \tilde{E}(\vec{r}')$$

$$= \left(\frac{\vec{\sigma}_p}{\sigma_E}\right) F_\alpha(\vec{r}')$$

subst. $r'_\alpha = \frac{r_\alpha - \mu_p}{\sigma_p} \rightarrow \frac{\partial r'_\alpha}{\partial r_\alpha} = \frac{1}{\vec{\sigma}_p}$

$\Rightarrow \; t(\vec{F}) \overset{!}{=} \left(\frac{\vec{\sigma}_P}{\sigma_E}\right) \vec{F}$ such that $(\mp)$ holds

$\Rightarrow$ everything is dimensionless!

$\Rightarrow$ independent of the original units but the thumb rule derived above is not directly applicable

- Note that $\vec{\sigma}_p \cdot \vec{F}$ means elementwise multiplication, e.g.

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$ & same when dividing