

Biomedical Data Science 2025: Homework Assignment 2

Due: Apr 30th (Wednesday) 11:59pm EST

Choose to do either MCDB & MBB (non-programming) or CBB & CS & S&DS (programming) assignment, depending on your academic affiliation. No late submissions will be accepted. Submission should be done in Canvas.

CBB & CPSC & S&DS (programming)

This year's programming assignment is given in an .ipynb file, which you can find [here](#) or on the class website. We highly recommend you run this on Google Colab. To do this:

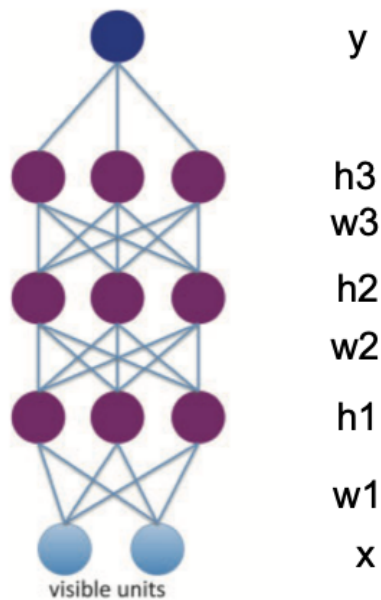
- Download the .ipynb file above for this assignment
- Go to <https://colab.research.google.com>
- Upload the .ipynb for this assignment
- Further instructions can be found in the notebook once you upload and start working on it
- When you are finished, submit both the ipynb and pdf output files on Canvas.

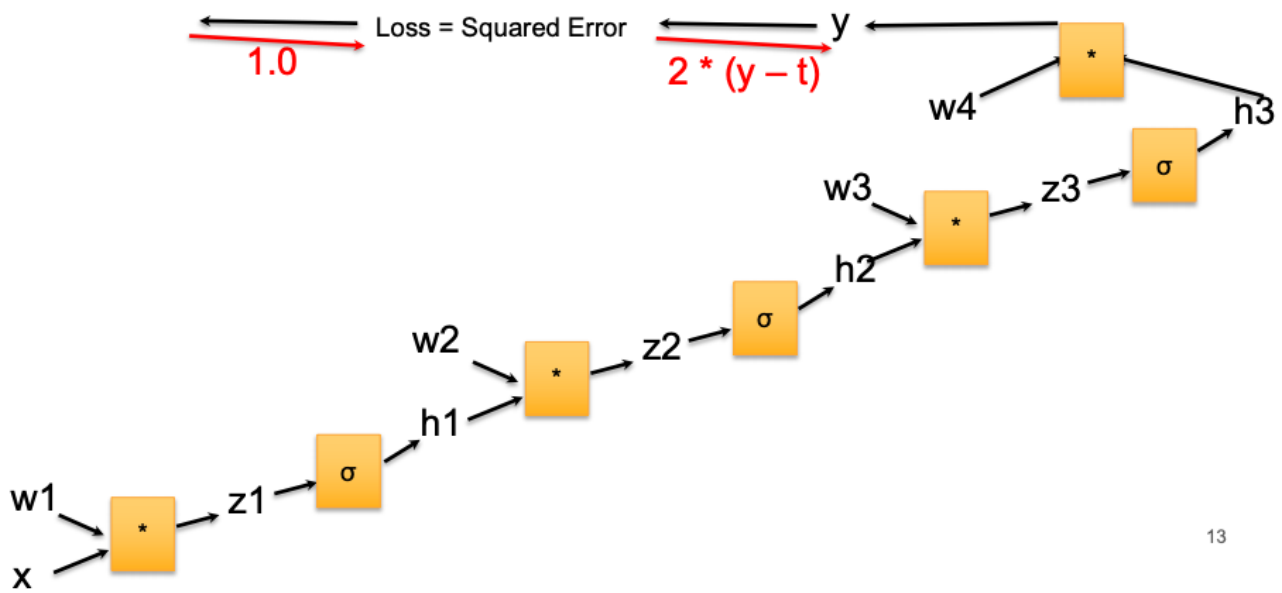
MBB&MCDB (non-programming)

Submit a single file answering the following questions.

1. (35pt) Derive the expressions for the x-, y-, and z-components of the force \vec{F}_j on atom $j=i+1$ from the previous atom i and successive atom $k=i+2$ using the bond angle potential, $V_{ba} = \frac{k_\theta}{2}(\theta_{ijk} - \theta_0)^2$, where k_θ is the constant bond stiffness, $\theta_{ijk} = \cos^{-1}\left(\frac{\vec{r}_{ij} \cdot \vec{r}_{kj}}{r_{ij}r_{kj}}\right)$ is the bond angle between bonded atoms i, j , and k , $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$, and θ_0 is the preferred bond angle. Note that $\vec{F}_j = \frac{-dV_{ba}}{dx_j} \hat{x} + \frac{-dV_{ba}}{dy_j} \hat{y} + \frac{-dV_{ba}}{dz_j} \hat{z}$.

2. (35pt) With the computation graph below, derive the gradient of the network with a linear output unit and a squared error loss based on backpropagation (with respect to the weight w_1).





13

3. (50pt) Read the following paper and write a short summary:

Grønbech, Christopher Heje, et al. "scVAE: Variational auto-encoders for single-cell gene expression data." *Bioinformatics* 36.16 (2020): 4415-4422.

In your summary, please try to answer these questions:

- What do the authors want to achieve?
- What is the major advantage of using variational autoencoders compared to other methods (esp. traditional autoencoders)?
- How is the input data represented (i.e. what is provided to the model)?
- How do the authors design the likelihood function and what is the intuition behind it? What experiments do the authors perform to show the effectiveness of the model?