DSSL@CSU, Fall 2020
`https://dssl-csu.github.io/about`
Leading Discussants: Daniel Cooley and Mantautas Rimkus, Colorado State University.

Meeting 1, 9/23/2020. Scribed by Mantautas Rimkus.

1. Introduction (warm-up)

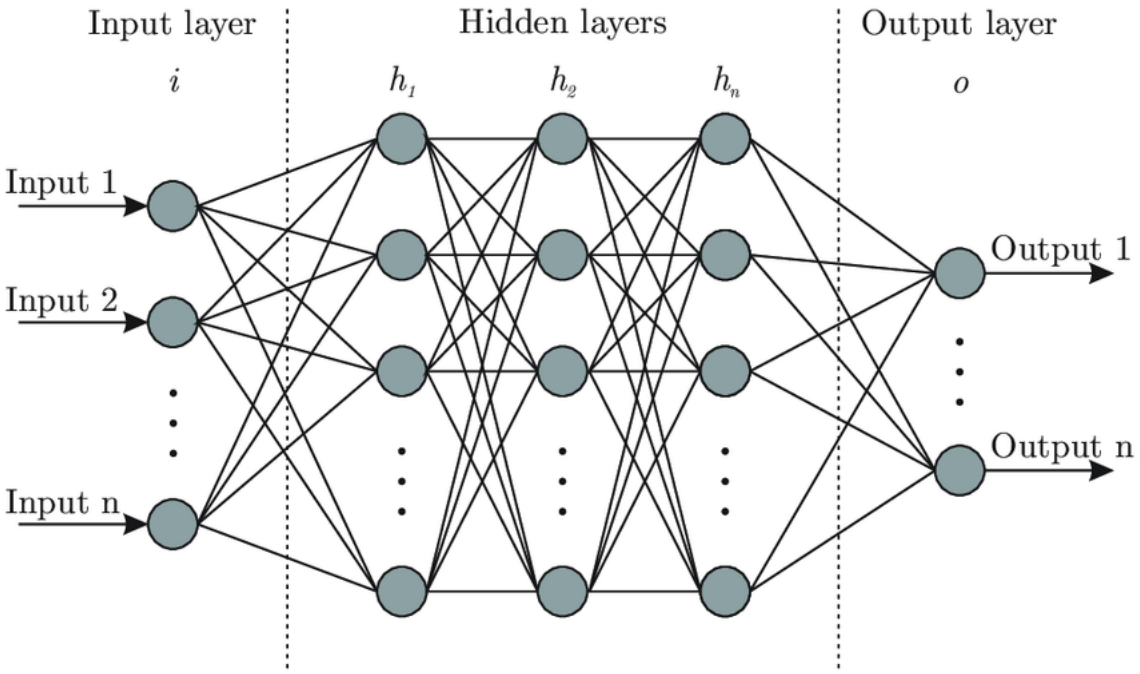

Figure 1: Neural Network

- Is anyone in Club tried any of methods that provide insights for AI methods?
- The main idea: "We show that we can decompose the pre-activation prediction values into a linear combination of training point activations, with the weights corresponding to what we call representer values, which can be used to measure the importance of each training point has on the learned parameter of the model."
- How you all understood this sentence: "a positive representer value indicates that a similarity to that training point is excitatory, while a negative representer value indicates that a similarity to that training point is inhibitory, to the prediction at the given test point".
- I think the authors are claiming in paragraph 3 that their "richer understanding" comes from the fact that their representer points can take on positive and negative values. Do other methods only return positive values?

2. Related Work:

    - Is anyone wants to share how much they are aware of the topic? Do anyone read any of these related works?
    - Should the authors' approach be lumped in with the second class: sample-based methods?

3. Representer Point Framework:

    - Examples of activation function σ?
    - Are these y vectors belonging to zero and one? Examples?
    - Explain me like I am five: what is the difference between Θ_1 and big Θ_2?
    - Can we interpret theorem 3.1 conditions as we must have L-2 to be able apply this theorem?
    - $\alpha_i j$ should have a large value, and $f_i^T f_t$ should have a large value" What is large?

4. 3.1 Training as Interpretable Model

    - It states that: "We emphasize that imposing L2 weight decay is a common practice to avoid overfitting for deep neural networks, which does not sacrifice accuracy while achieving a more interpretable model."

5. Pre-trained models

    - In general, I did not understand motivation behind this subchapter. What would be a real world examples that would motivate it? I agree that this is confusing. It seems to be talking about agreement with the activator function?
    - Figure1: why there are some points, that lays on left-top corner, or right-bottom (disagreements)?

6. Figure3 and Figure 4.

References

[1] Yeh, C-K., Kim, J.S., Yen, I.E.H., and Ravikumar, P. (2018). *Representer Point Selection for Explaining Deep Neural Networks. NIPS*