Deep Neural Regression Collapse

Akshay Rangamani, Altay Unal

New Jersey Institute of Technology, Newark, NJ, USA.
Contact: akshay.rangamani@njit.edu, au252@njit.edu

Main Contributions

Neural Regression Collapse (NRC) [1] is a recently identified phenomenon which characterizes the layer representations, also known as features, and model weights within deep networks trained on regression problems, similar to Neural Collapse [2] in classifiers. We investigate this phenomenon in the intermediate layers of deep regressors. We refer to this as **Deep Neural Regression Collapse** characterized by the following properties:

- **Noise Component Suppression**: A collapsed layer in a deep regressor only extracts the signal relevant to the target prediction and minimizes the noise
- **Feature-Weight Alignment:** The weights of a collapsed layer are low rank and equivalent to the signal component of the features at that layer.
- **Feature-Target Covariance Alignment:** The signal-target covariance aligns with the target covariance in a collapsed layer.

Measurements

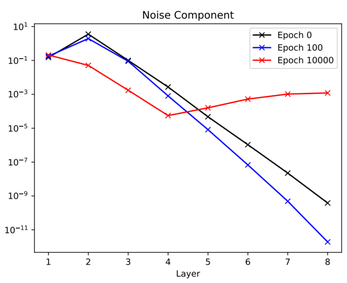
- **Noise Component Suppression**: We plot $Tr(\Sigma_{H^\ell}) - Tr(V^\top \Sigma_{H^\ell} V)$ through the layers to observe the noise component where $V \in R^{d \times t}$ denotes the top t eigenvectors of the feature covariance matrix for the ℓ-th layer Σ_{H^ℓ}. We also plot the stable rank of the feature matrix H^ℓ to show that it is low dimensional.
- **Feature-Weight Alignment:** We plot the average cosine of the angles between the input subspace of W_ℓ (weights) and the column space of H^ℓ (features). We also plot the stable rank of W_ℓ to show that the weights are also low rank.
- **Feature-Target Covariance Alignment :** We plot the average cosine of the angles between the relevant subspaces of the feature–target covariance $\Sigma_{H^\ell Y}$ and the target covariance Σ_Y. For targets that are rank-t this is trivially 1, but for low rank targets, we can investigate whether the deep network learns the low dimensional subspace.

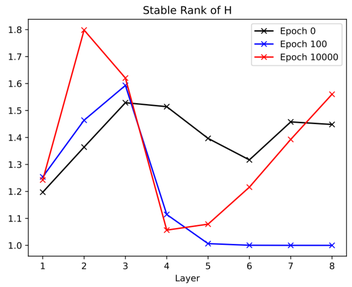
Observations

- We report the measurements for noise component suppression and feature-weight alignment for networks trained on imitation learning in the reacher and swimmer MuJoCo environments. Both conditions of collapse are satisfied
- We measure the rank of the weights and observe that collapsed layers are low rank. We also measure the stable rank of the features and observe that the inputs are first projected into a high-dimensional space, then in subsequent layers, the features are projected into lower dimensional subspaces corresponding to the target dimension.
- To observe the feature-target covariance alignment in a non-trivial setting, we train a deep network on low rank target data generated synthetically. We find that for targets that lie in a 2-dimensional subspace of $\mathbb{R}^{10}$, the alignment between the top 2-dimensional subspaces of $\Sigma_{H^\ell Y}$ and Σ_Y is perfect, while the top 2-dimensional feature subspace is orthogonal to the remaining 8-dimensional subspace of the target. We can also observe this by measuring the feature-weight alignment.

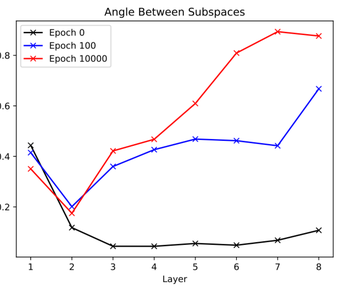
Results

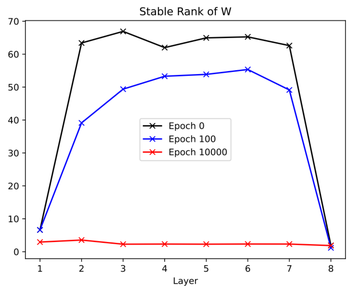
Noise Component Suppression



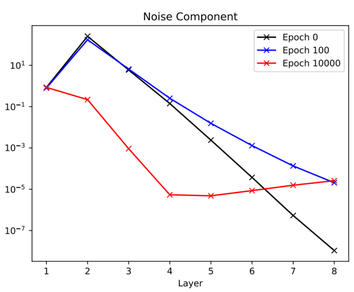


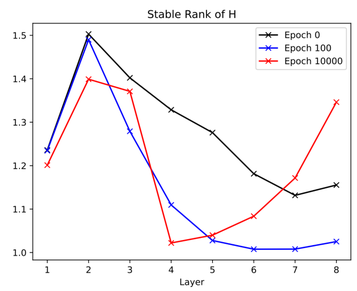
Feature – Weight Alignment

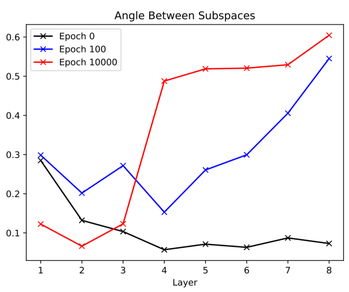


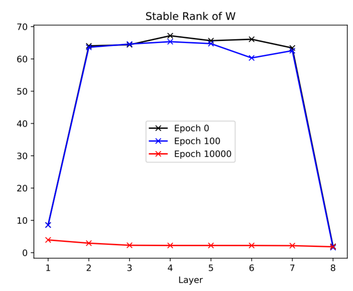


Imitation Learning on Swimmer - 8 layer MLP with 256 hidden nodes



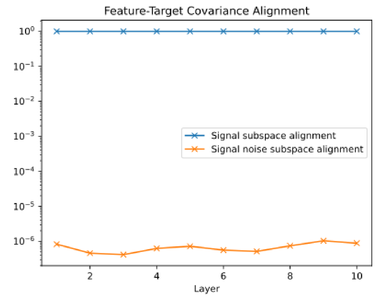


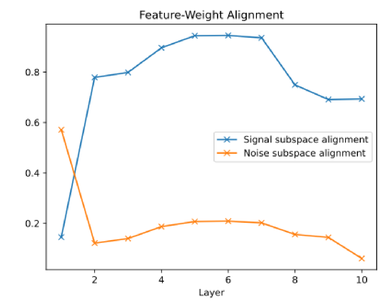




Imitation Learning on Reacher - 8 layer MLP with 256 hidden nodes

Low Rank Targets: Feature – Target Covariance Alignment





Synthetic data experiment – 10 layer MLP with 512 hidden nodes, Target ∈ rank 2-subspace of $\mathbb{R}^{10}$

Conclusions

- Neural Regression Collapse can be observed not just in the top layer of deep regressors, but also in the deeper in the network
- There is a correspondence between the rank of the learned weights, the rank of the features, and the rank of the target
- Neural Collapse describes not just deep classifiers but also deep regression models.

References

1. Andriopoulos, G., Dong, Z., Guo, L., Zhao, Z., & Ross, K. (2024). The prevalence of neural collapse in neural multivariate regression. *arXiv preprint arXiv:2409.04180*.
2. Papyan, V., Han, X. Y., & Donoho, D. L. (2020). Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, *117*(40), 24652-24663.
3. Rangamani, A., Lindegaard, M., Galanti, T., & Poggio, T. A. (2023, July). Feature learning in deep classifiers through intermediate neural collapse. In *International Conference on Machine Learning* (pp. 28729-28745). PMLR.
4. Súkeník, P., Mondelli, M., & Lampert, C. H. (2024). Deep neural collapse is provably optimal for the deep unconstrained features model. *Advances in Neural Information Processing Systems*, *36*.