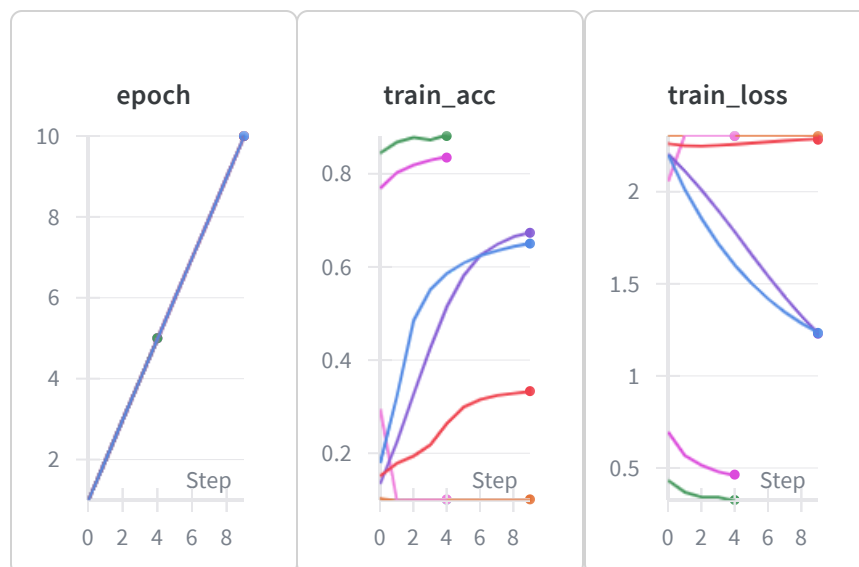


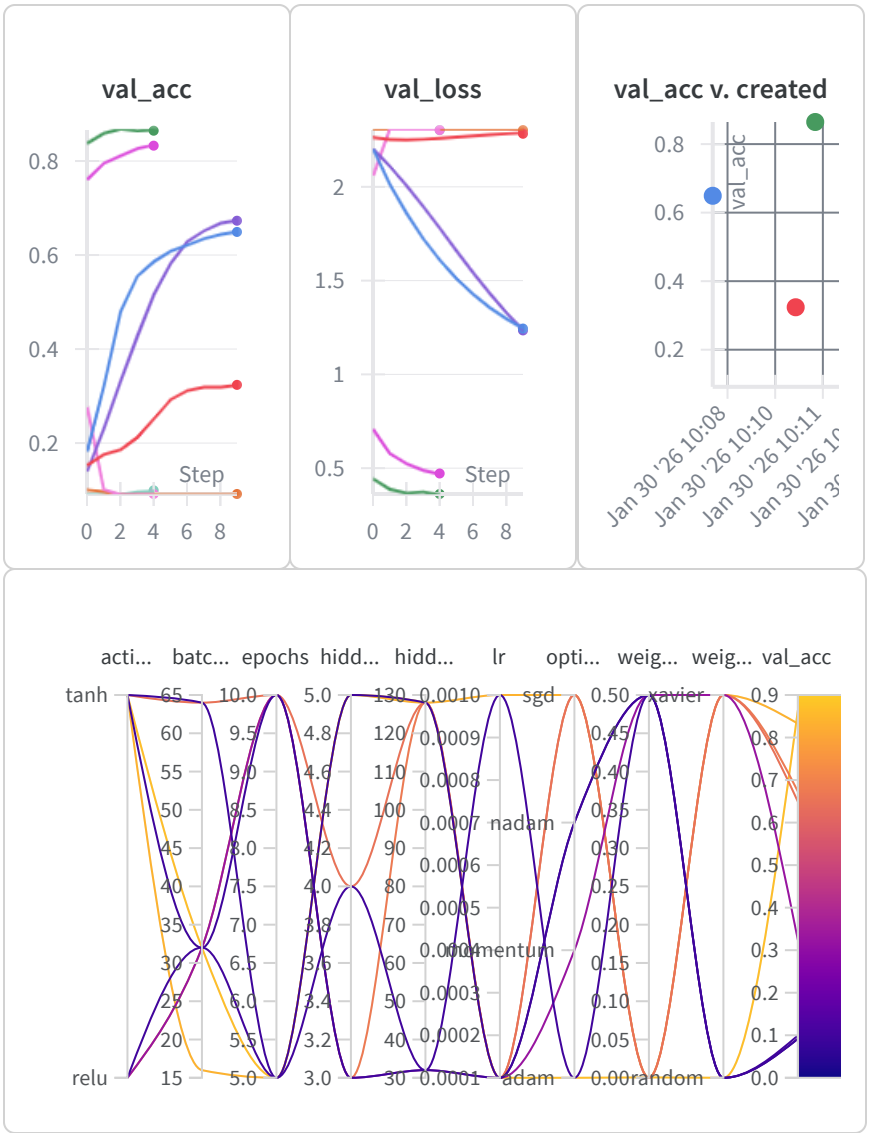
Hyperparameter Tuning and Model Insights

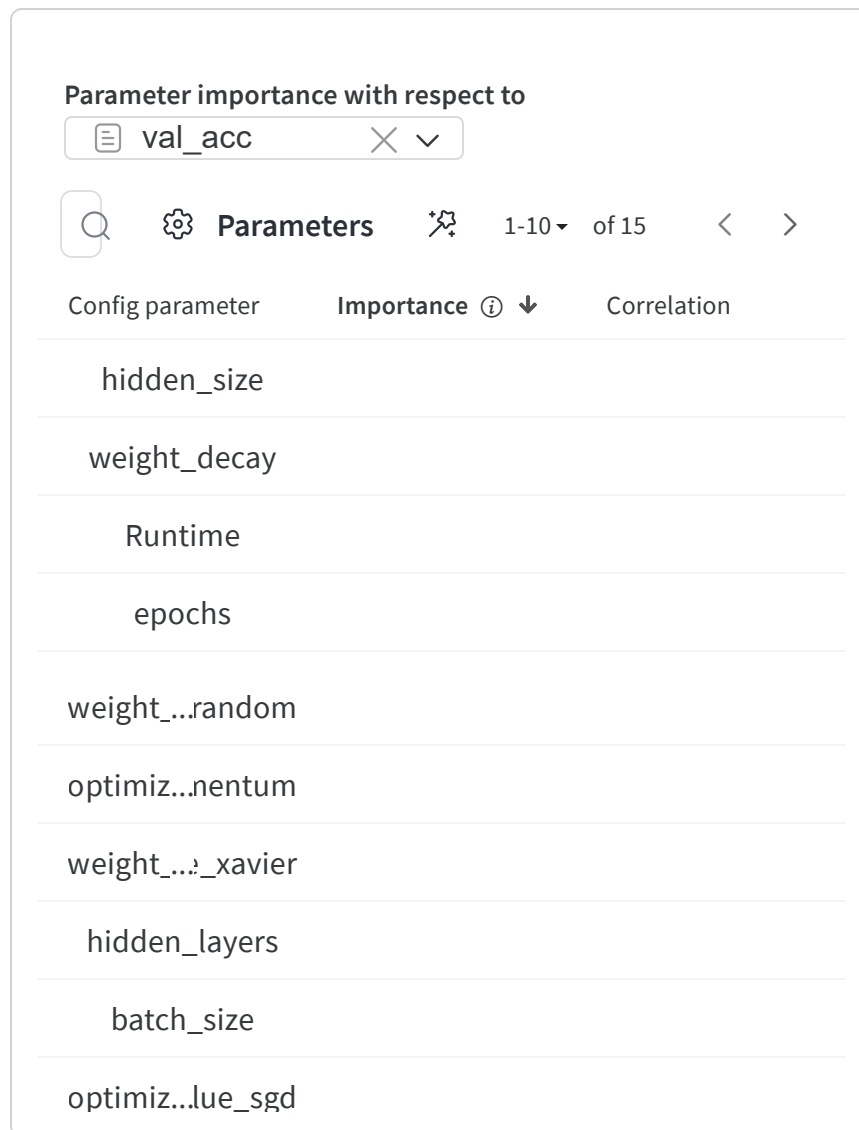
Chosen strategy is Bayesian search. Grid search is not suitable here because we have many hyperparameters, and evaluating all combinations would cause the search space to grow exponentially, requiring thousands of runs. Random search might work but Bayesian search builds a probabilistic model to find and explore deeper into regions that offers more promising results while reducing the number of runs needed to identify the high-performing configurations, and that's why I chose Bayesian search strategy.

Chellappan Chellappan

Created on January 30 | Last edited on January 30







Observations derived from reading the Parallel Co-ordinates Plot and Correlation Summary:

- Weight decay has lowest correlation and configurations with higher weight decays are resulting in lower accuracies.
- Tanh activation consistently achieves higher accuracy than ReLU
- Unlike what I expected, batch size shows very minimal correlation.
- Configurations with higher hidden sizes like 128 are producing results with higher accuracies.
- While testing on my own, sgd produced better results than nadam or momentum optimizers and the plot has confirmed it for me.
- Lower learning lead to more stable and higher accuracies whereas higher learning rate is introducing instability.

- Hidden layer size has the strongest influence on the accuracy levels based on the correlation summary.

Created with  on Weights & Biases.

<https://wandb.ai/cl-chellappan-atri-ai/Fashion%20MNIST%20Hyperparameter%20Tuning%20with%20wandb/reports/Hyperparameter-Tuning-and-Model-Insights---VmldzoxNTc5NzQ1Mg>