Machine Learning Lessons: Reflections on a Month of Practice

Practical Insights on Readmes, MIGs, and Daily Movement in Machine Learning

Leveraging README Documents for Enhanced Personal Machine Learning Project Management

- Create README documents for your future self to document project details.

- Include quickstart instructions, data locations, reproduction steps, and common pitfalls.

- A well-maintained README saves time reconstructing past work, especially after project pauses.

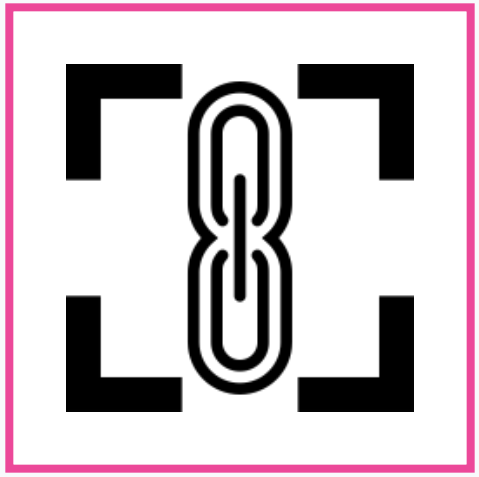
Optimizing Machine Learning Resource Allocation with Multi-Instance GPU Slices

- Request Multi-Instance GPU (MIG) slices instead of full GPUs for smaller models.

- MIG allows partitioning a single GPU into smaller, isolated virtual GPUs.

- Using MIG slices can lead to faster job scheduling and improved iteration speed.

Incorporating Movement Snacks to Combat Prolonged Sitting in Machine Learning

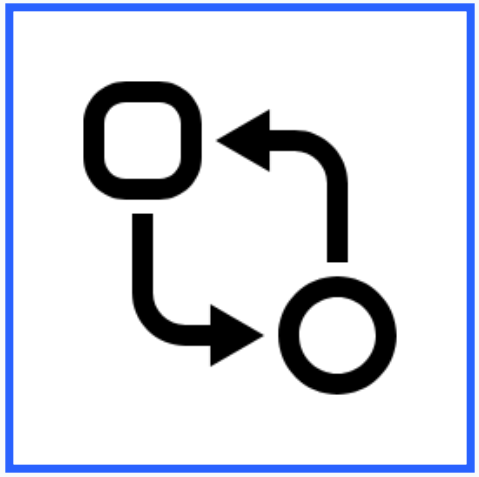


- Alternate positions and incorporate short movement breaks to combat poor posture.

- Include exercises like band pull-aparts, wall stretches, and hip flexor stretches.

- Consider standing during audio meetings or reading blocks to promote better posture.

Documenting Environment Setup for Reproducible Machine Learning Experiments



- Specify the exact Python version and environment file or conda/pip commands used.

- This ensures that the environment can be easily recreated for reproducibility.

- Include specific versions of key libraries to avoid compatibility issues.

Tracking Data Provenance and Processing Steps in Machine Learning Projects

- Record the locations of raw and processed data, including download and caching details.

- Note any data-specific quirks or transformations applied during preprocessing.

- Use checksums to verify data integrity and prevent data corruption.

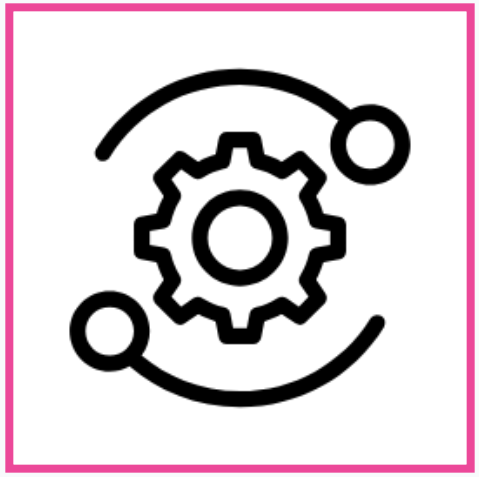
Creating Reproducible Results by Documenting Training and Evaluation Commands

- Provide the exact commands used to run the main experiments, including all flags.

- Document how to resume training from checkpoints and how to set random seeds.

- This ensures that the same results can be obtained consistently.

Logging Hyperparameter Search Configurations and Results for Machine Learning



- Record the command used for hyperparameter search, including the search ranges.

- Specify where the results of the hyperparameter search are logged.

- This allows you to easily review and compare different hyperparameter configurations.

Maintaining a Changelog to Track Meaningful Changes in Machine Learning Projects

- Include one-line bullets of meaningful changes made to the project.

- This helps track the evolution of the project and identify the source of issues.

- The changelog should include bug fixes, feature additions, and performance improvements.