

Benchmark: MMLU Machine Learning
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

