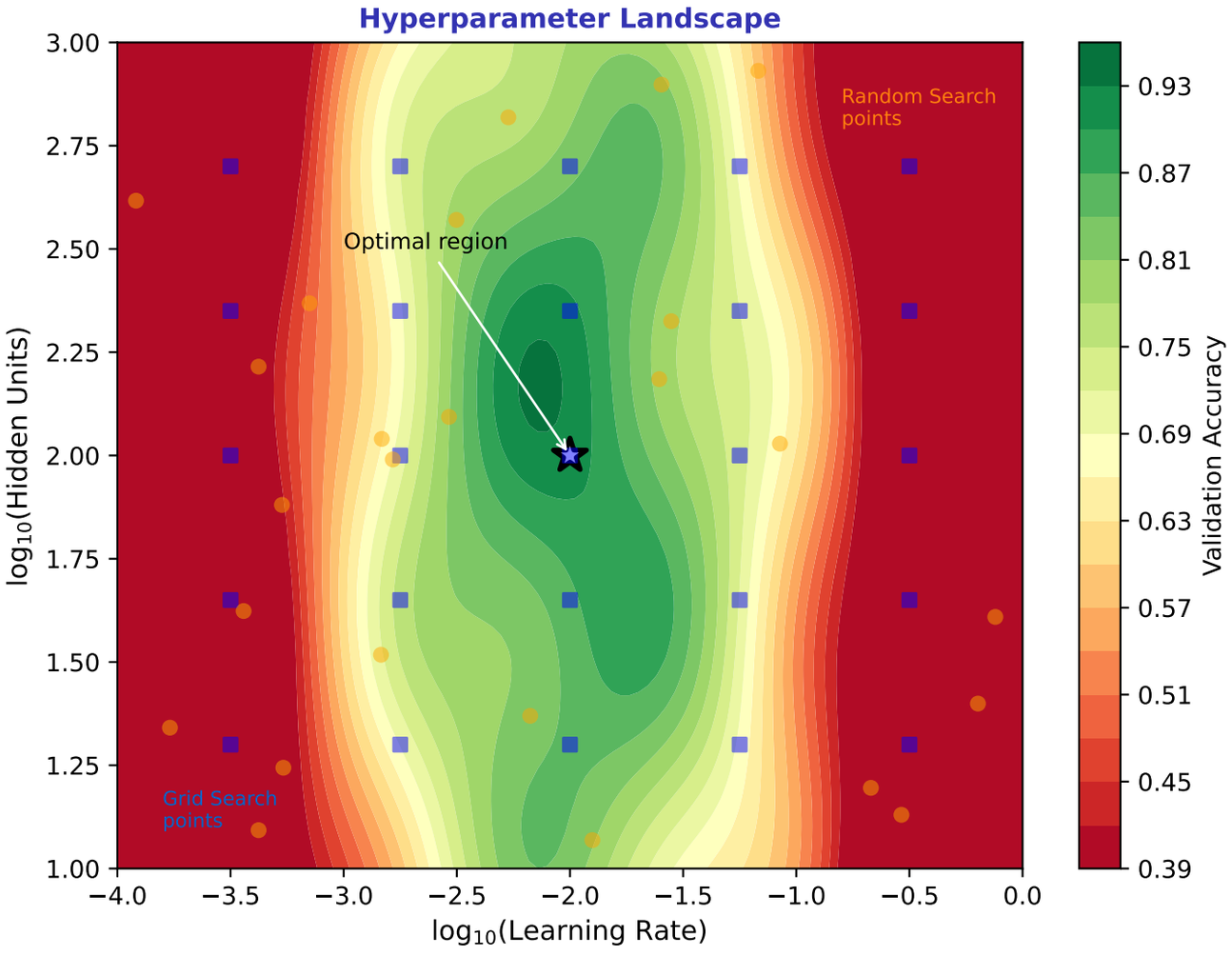
Hyperparameter Optimization: Finding the Best Configuration



Hyperparameter Search Methods

Grid Search

- Exhaustive search over grid
- Exponentially expensive
- May miss optimal region

Random Search

- Sample randomly from ranges
- Often better than grid
- More efficient exploration

Bayesian Optimization

- Uses prior results
- Smart exploration
- Best for expensive evaluations

Key Hyperparameters to Tune:
- Learning rate: 1e-4 to 1e-1
- Hidden layer size: 32 to 1024
- Number of layers: 1 to 5
- Batch size: 16 to 256
- Regularization strength: 1e-5 to 1e-1
- Dropout rate: 0.1 to 0.5