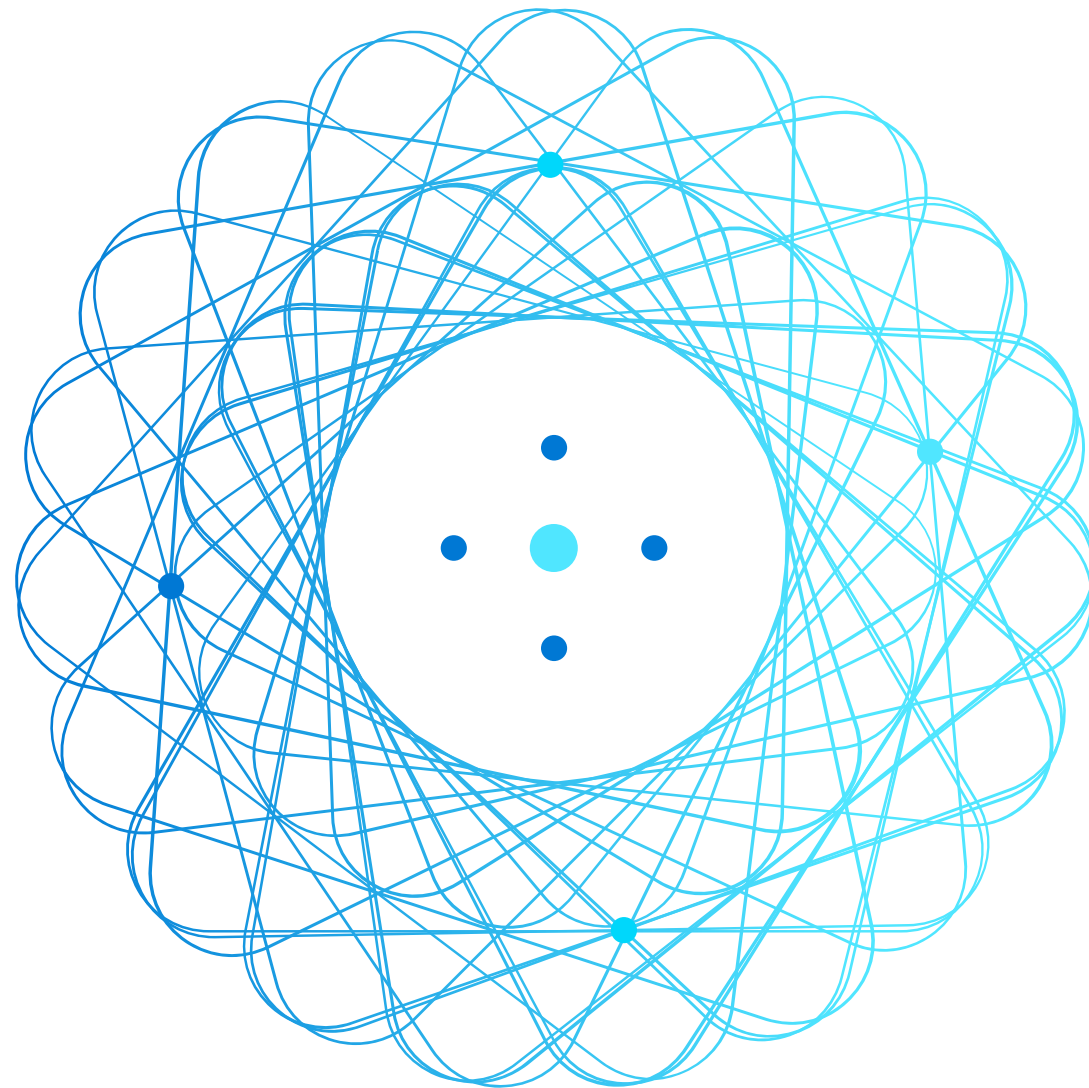


Module 8: Training Optimal Models

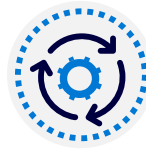
Mohammed Arif



Agenda

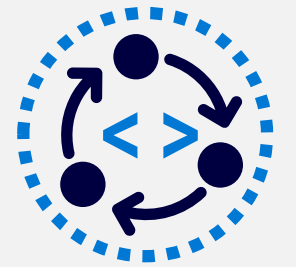


Hyperparameter Tuning



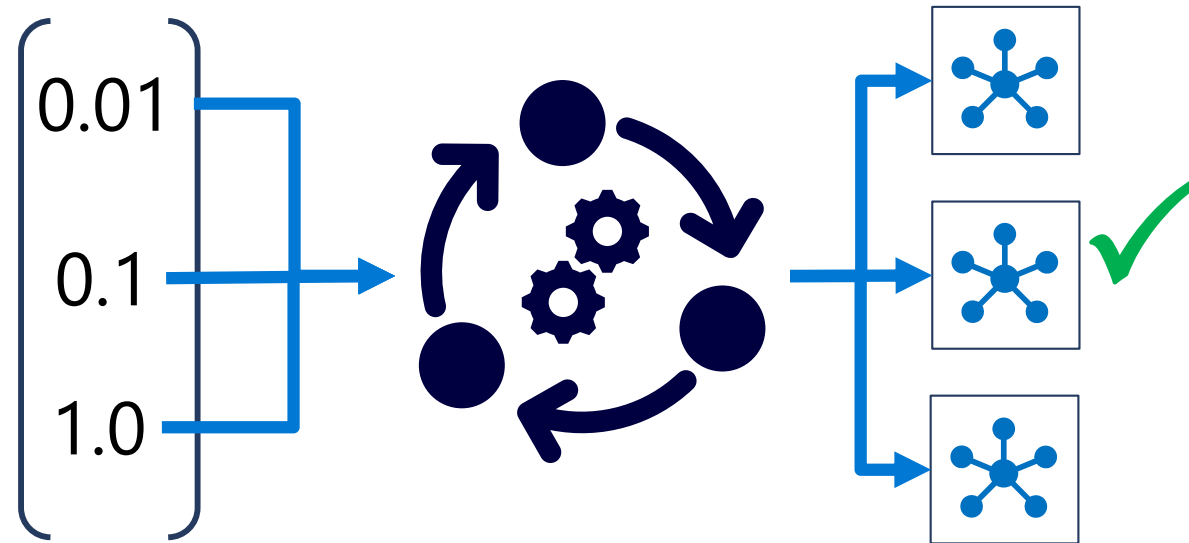
Automated Machine Learning

Hyperparameter Tuning



What is Hyperparameter Tuning?

Train multiple models, using the same algorithm but varying hyperparameter values
Find the "best" model based on a specific performance metric



Hyperparameter Search Space

Discrete Hyperparameters

Choice (any list or range)

From a discrete distribution (qnormal, quniform, qlognormal, qloguniform)

Continuous Hyperparameters

From a continuous distribution (normal, uniform, lognormal, loguniform)

```
param_space = {  
    '--batch_size': choice(16, 32, 64),  
    '--learning_rate': normal(10, 3)  
}
```

Hyperparameter Sampling

Grid Sampling

Tries every combination of discrete hyperparameter values

Can only be used when all hyperparameters are discrete

Random Sampling

Randomly selects hyperparameter values

Can be used with discrete and continuous hyperparameter combinations

Bayesian Sampling

Selects hyperparameter values based on performance of previous selection

Can only be used with **choice**, **uniform**, and **quniform** hyperparameters

```
from azureml.train.hyperdrive import RandomParameterSampling

param_sampling = RandomParameterSampling(param_space)
```


Tuning Hyperparameters with Hyperdrive

Experiment script

```
parser.add_argument('--reg', type=float, dest='reg_rate')  
...  
run.log('Accuracy', model_accuracy)
```

Hyperparameters in sampling collection are passed as arguments

Log performance metric for evaluation

Hyperdrive run configuration

```
hyperdrive = HyperDriveConfig(run_config=script_config,  
                              hyperparameter_sampling=param_sampling,  
                              policy=stop_policy,  
                              primary_metric_name='Accuracy',  
                              primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,  
                              max_total_runs=6,  
                              max_concurrent_runs=4)
```

ScriptRunConfig for training script

Params added to script arguments

Name must match logged metric

```
hyperdrive_run = experiment.submit(config=hyperdrive)
```


Lab: Tune Hyperparameters



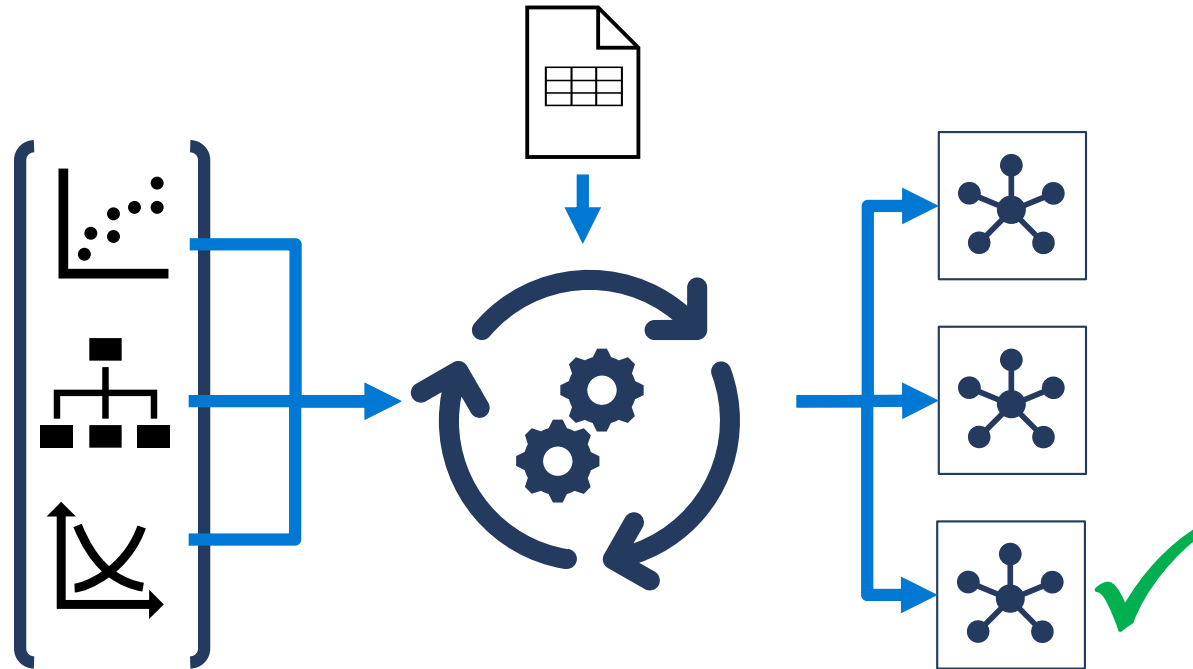
1. View the lab instructions at <https://aka.ms/mslearn-dp100>
2. Complete the **Tune hyperparameters** exercise

Automated Machine Learning



Automated Machine Learning – A Reminder

Train multiple models in parallel, varying algorithm and preprocessing
Find the "best" model based on a specific performance metric



Preparing Data for Automated Machine Learning

Training Data – tabular data including features and label

Validation Data – optional table for model validation

You can use a **Dataset**
or a Pandas dataframe

```
tab_ds = ws.datasets.get("tabular dataset")  
  
train_ds, test_ds = tab_ds.random_split(percentage=0.7, seed=123)
```

Optional split for training and test
(if only training data is provided, cross-validation
will be applied automatically)

Running an Automated Machine Learning Experiment

Configure an automated machine learning experiment run

```
from azureml.train.automl import AutoMLConfig

automl_config = AutoMLConfig(name='Automated ML Experiment',
                             task='classification',
                             compute_target=aml_cluster,
                             training_data = train_ds,
                             validation_data = test_ds,
                             label_column_name='Label',
                             iterations=20,
                             primary_metric = 'AUC_weighted',
                             max_concurrent_iterations=4,
                             featurization='auto')

automl_run = automl_experiment.submit(automl_config)
```

Metrics are dependent on task

(use `automl_utils.get_primary_metrics`
to find them)

Monitoring and Reviewing Automated ML Runs

Monitor runs in Azure Machine Learning studio or widget

Find the best-performing run and the model it trained:

```
best_run, fitted_model = automl_run.get_output()
best_run_metrics = best_run.get_metrics()
for metric_name in best_run_metrics:
    metric = best_run_metrics[metric_name]
    print(metric_name, metric)
```

View model pipeline details:

```
for step_ in fitted_model.named_steps:
    print(step_)
```

Lab: Use Automated Machine Learning from the SDK




1. View the lab instructions at <https://aka.ms/mslearn-dp100>
2. Complete the **Use automated machine learning from the SDK** exercise

Knowledge check

 You want to try every possible combination of a set of specified discrete values in a hyperparameter tuning experiment. Which kind of sampling should you use?

- ☒ Grid Sampling
 - ☐ Random Sampling
 - ☐ Bayesian Sampling
-

 You want to use automated machine learning to find the model with the best *AUC_weighted* metric. Which parameter of the *AutoMLConfig* object should you set?

- ☐ task='AUC_weighted'
- ☐ label_column_name= 'AUC_weighted'
- ☒ primary_metric='AUC_weighted'

References

Microsoft Learn: Tune hyperparameters with Azure Machine Learning

<https://docs.microsoft.com/learn/modules/tune-hyperparameters-with-azure-machine-learning/>

Microsoft Learn: Automate machine learning model selection with Azure Machine Learning

<https://docs.microsoft.com/learn/modules/automate-model-selection-with-azure-automl/>

Azure Machine Learning hyperparameter tuning documentation

<https://docs.microsoft.com/azure/machine-learning/how-to-tune-hyperparameters>

Azure Machine Learning automated machine learning documentation

<https://docs.microsoft.com/azure/machine-learning/how-to-configure-auto-train>

