



DEV
DAY

통합 머신러닝 플랫폼 Amazon SageMaker 활용하기

핸즈온랩

강지양 / AWS Solutions Architect
김태현 / AWS Solutions Architect





DEV
DAY

실습 자료 다운로드

<http://bit.ly/sagemaker11> (실습가이드)

<http://bit.ly/sagemaker11s> (슬라이드)

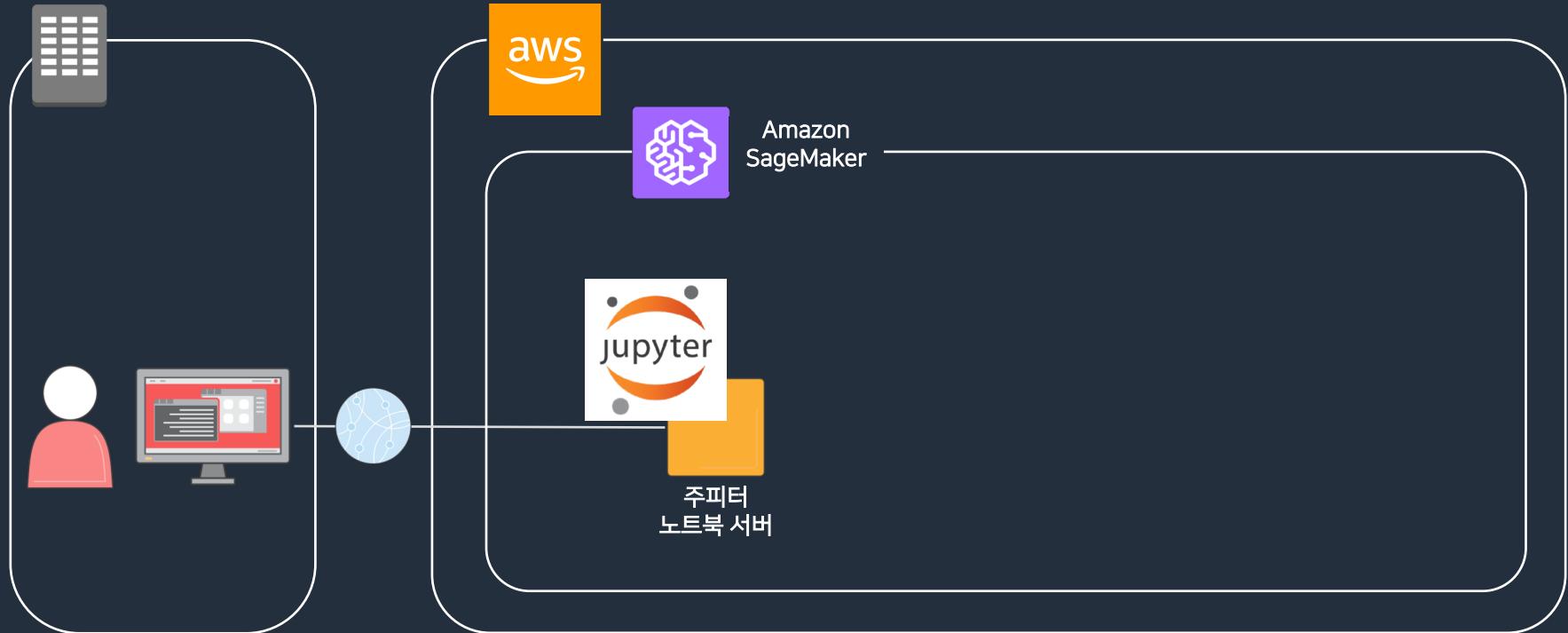


Amazon SageMaker

데이터 과학자와 개발자들이
머신러닝 모델을
빠르고 쉽게 만들도록 해주는
완전 관리형 서비스

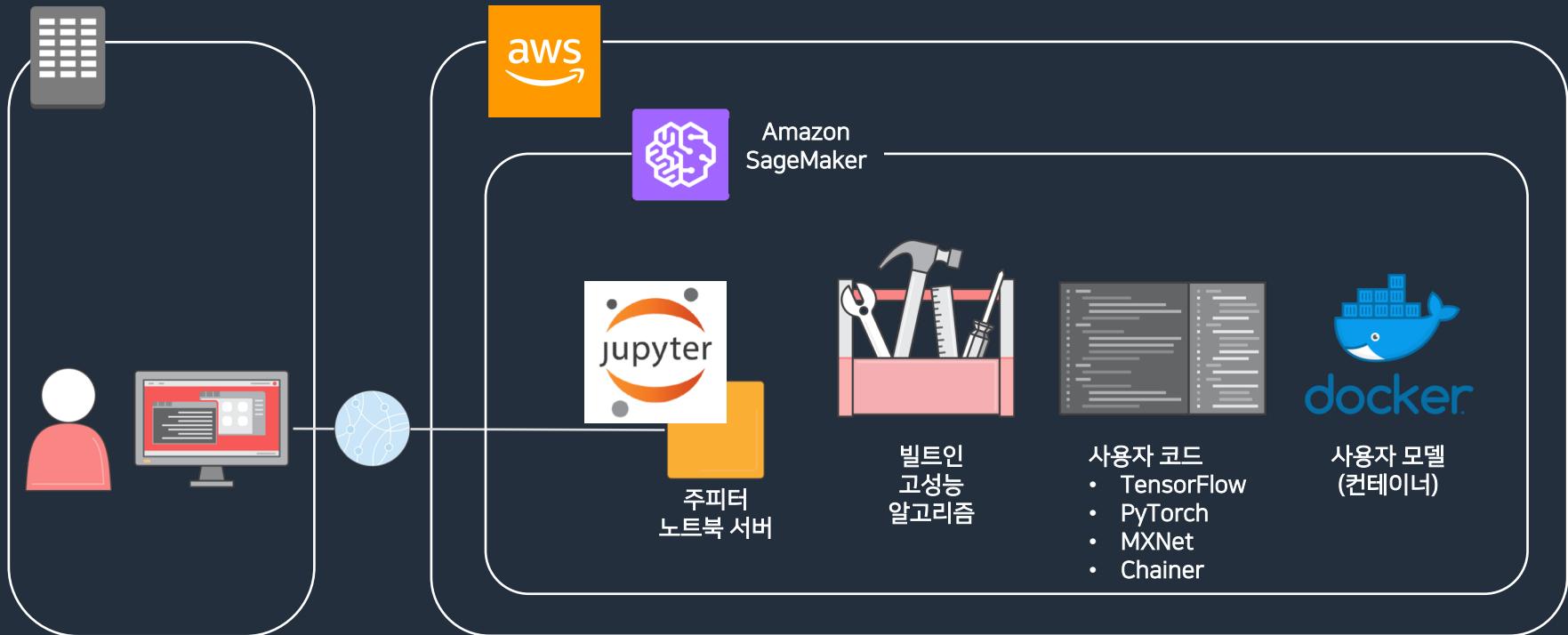
SageMaker란?

1. 주피터 노트북 개발 환경



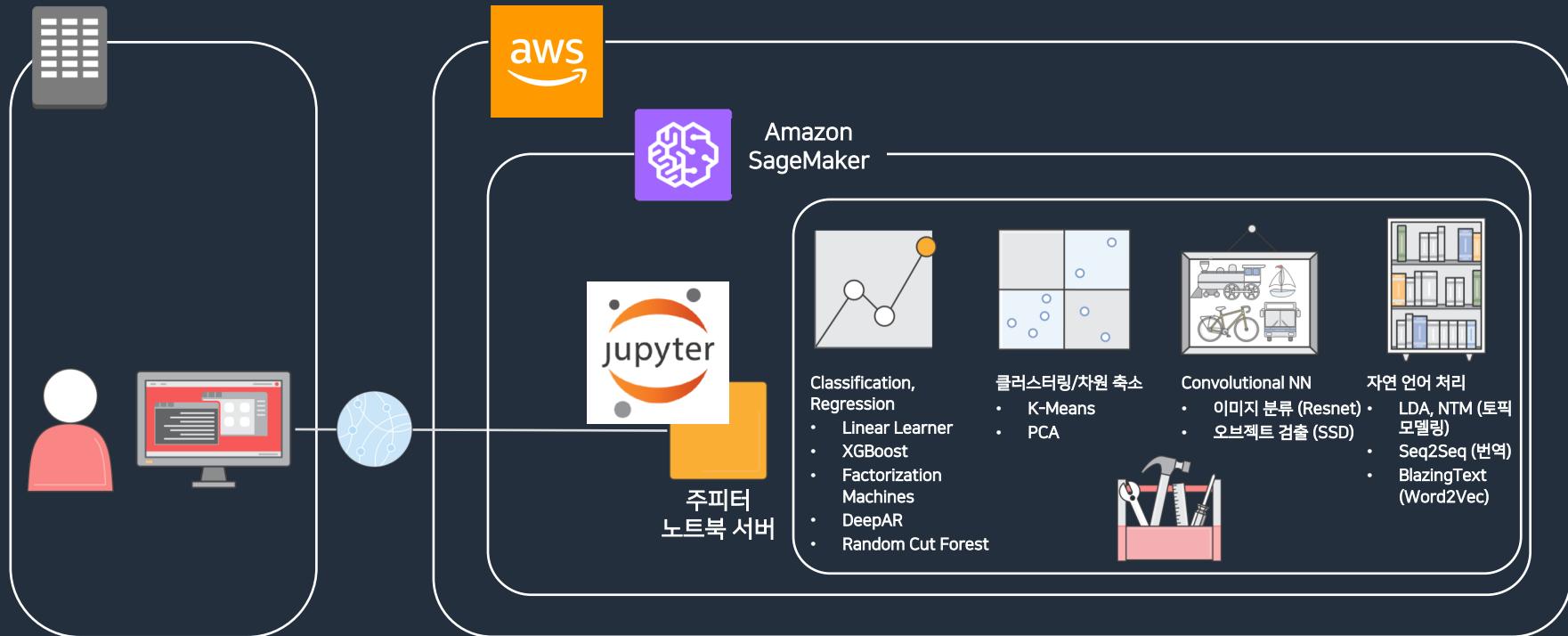
SageMaker란?

1. 주피터 노트북 개발 환경



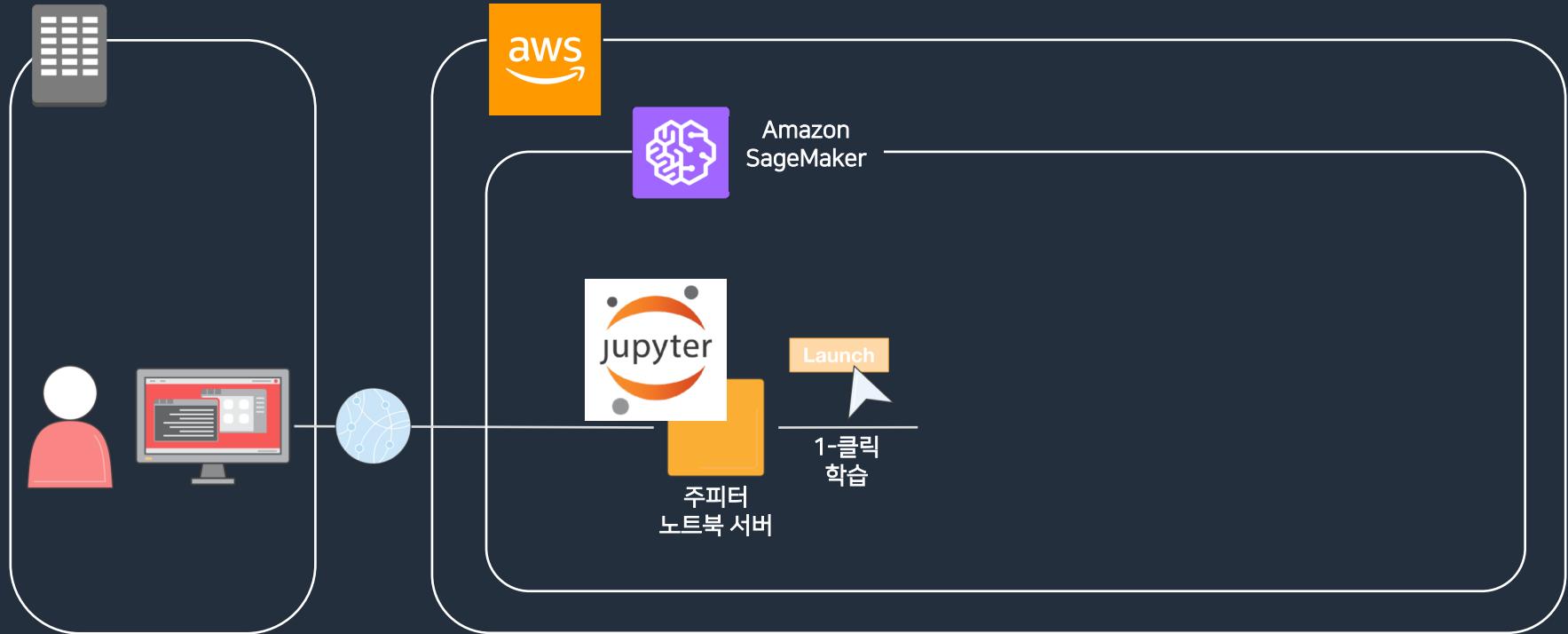
SageMaker란?

2. 빌트인 고성능 알고리즘



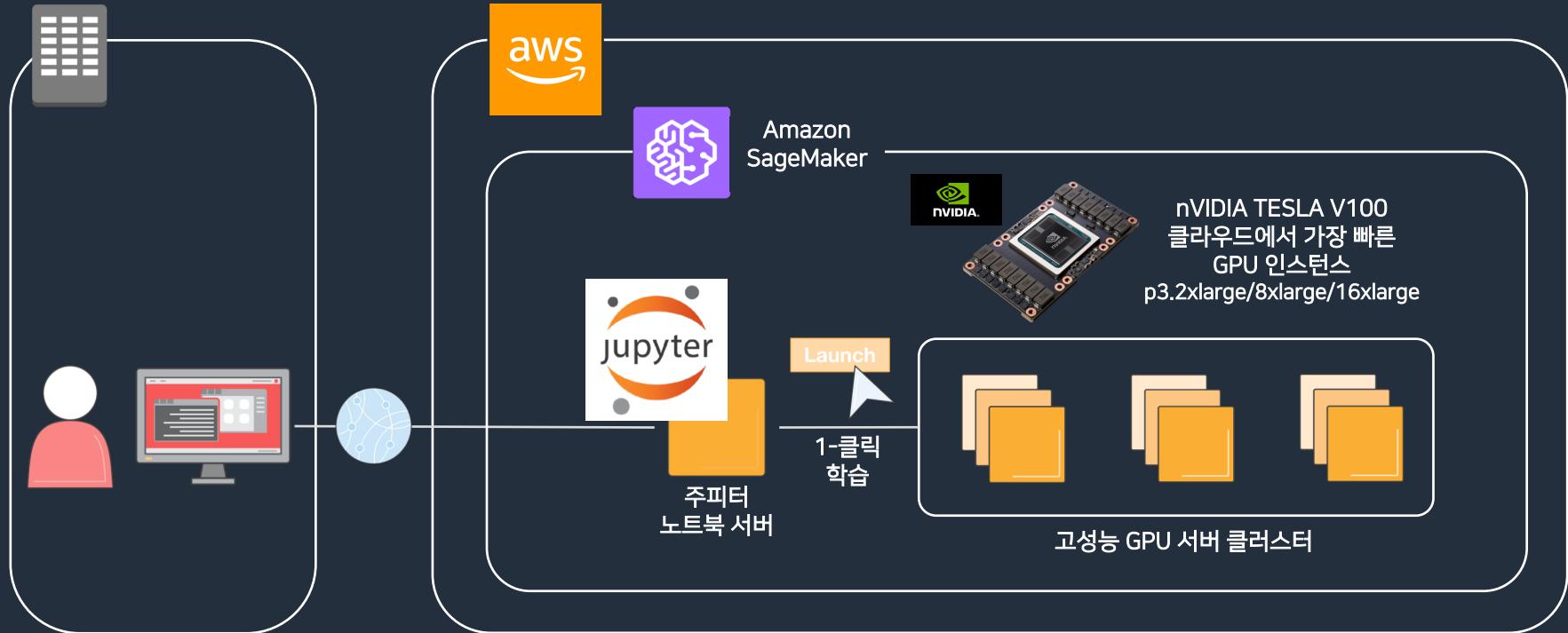
SageMaker란?

3. 분산 학습 서비스



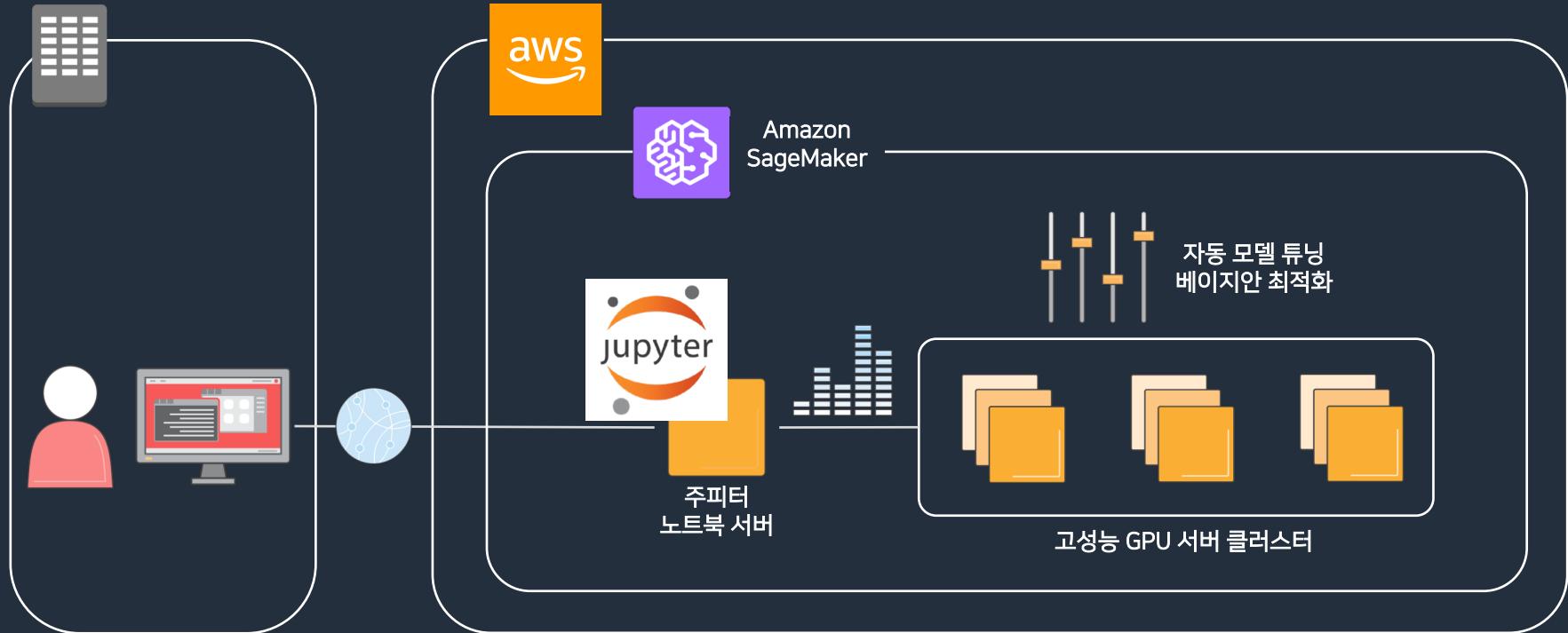
SageMaker란?

3. 분산 학습 서비스

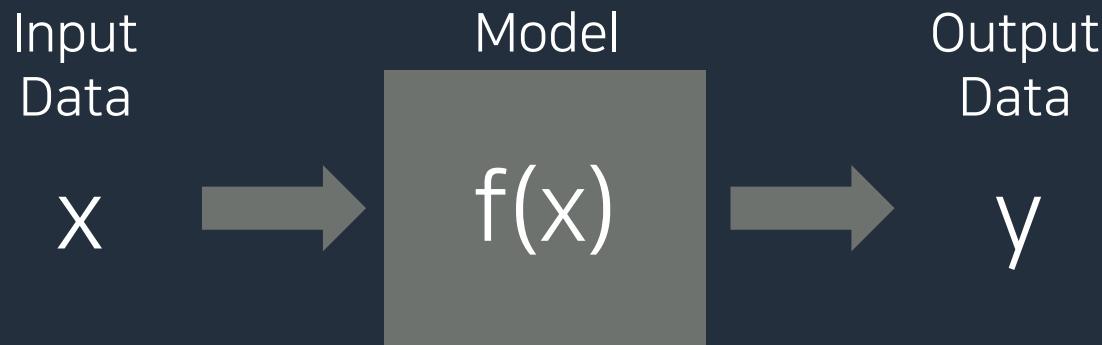


SageMaker란?

3. 분산 학습 서비스



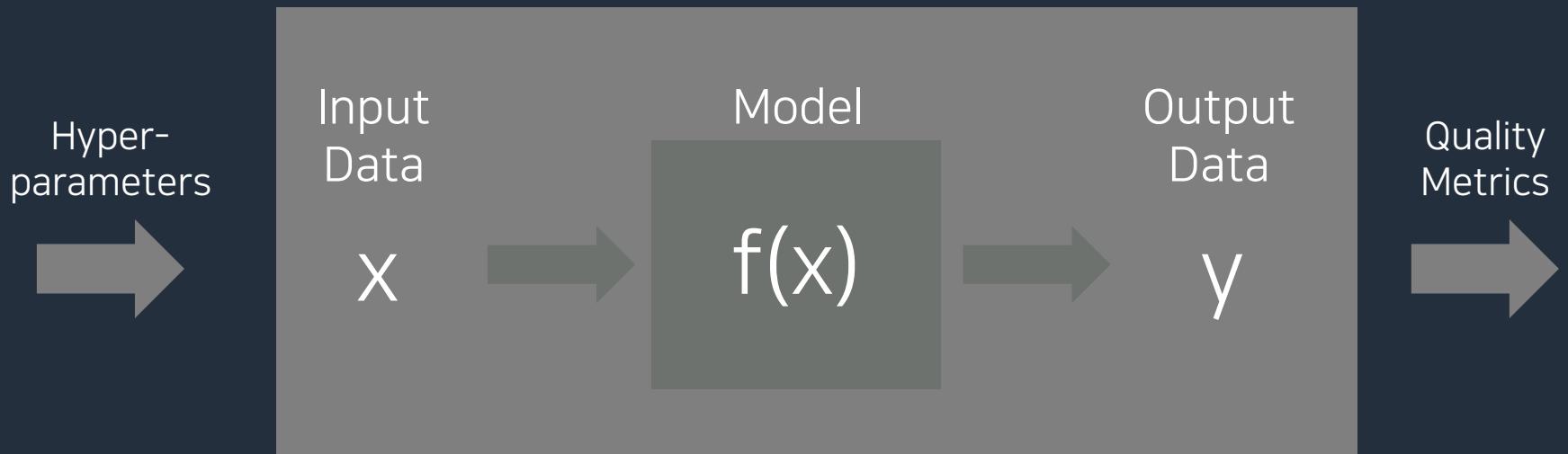
Machine Learning



for a given hyperparameter set

Meta-Machine Learning

Automatic Model Tuning



HPO: Theory

- "Practical Bayesian Optimization of Machine Learning Algorithms" by Jasper Snoek, Hugo Larochelle, and Ryan P. Adams (NIPS 2012)
- <https://youtu.be/MnHCe8tGjQ8>
(Review in Korean)

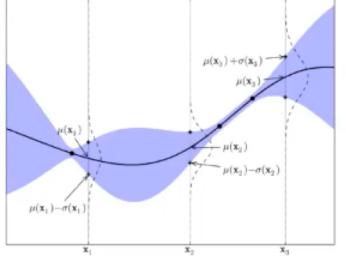
Practical Bayesian Optimization of Machine Learning Algorithms

NIPS 2012

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams

[PR12] PR-081
Apr. 15, 2018

Paper Review by Jiyang Kang



0:03 / 39:08

PR-080: Practical Bayesian Optimization of Machine Learning Algorithms

340 views

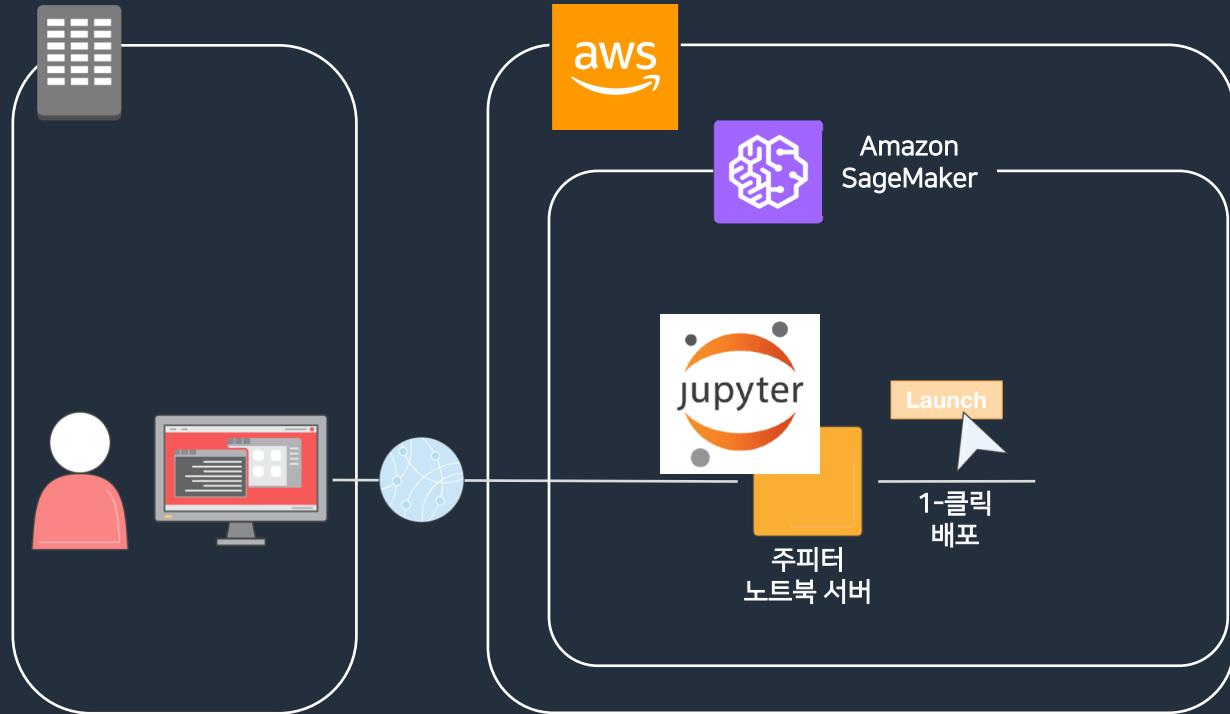
Jiyang Kang
Published on Apr 15, 2018

PAPER REVIEW BY JIYANG KANG

Paper review: "Practical Bayesian Optimization of Machine Learning Algorithms" by Jasper Snoek, Hugo Larochelle, and Ryan P. Adams (NIPS 2012)

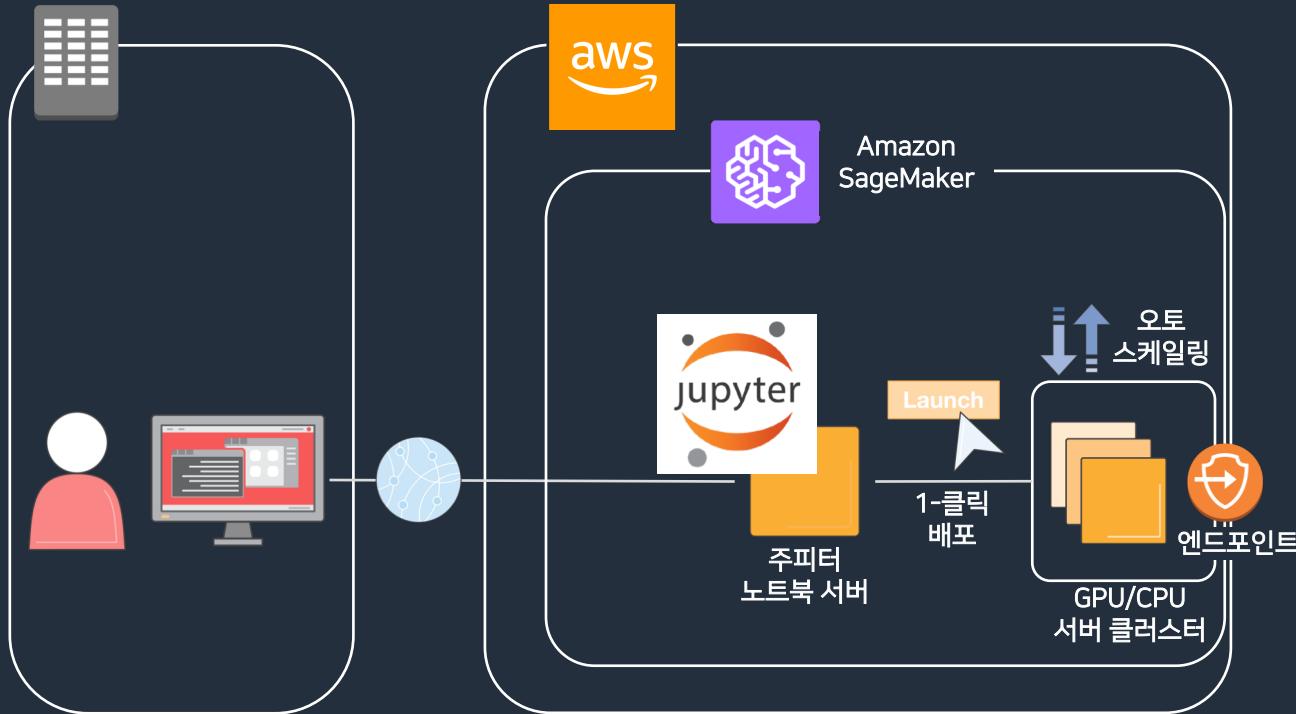
SageMaker란?

4. 예측 모델 배포 서비스



SageMaker란?

4. 예측 모델 배포 서비스



SageMaker란?

4. 예측 모델 배포 서비스



핸즈온 소개

핸즈온

1. Factorization Machine (빌트인 알고리즘 사용 예제)
2. MNIST (TensorFlow 사용 예제)
3. MNIST (TensorFlow 자동 모델 튜닝 예제)
4. HPO Analyzer (자동 모델 튜닝 결과 분석 예제)

1. 빌트인 알고리즘 사용하기

알고리즘 지정

```
from sagemaker.amazon.amazon_estimator import get_image_uri
container = get_image_uri(boto3.Session().region_name, 'factorization-machines')
```

인스턴스 설정

```
import boto3
import sagemaker

sess = sagemaker.Session()

fm = sagemaker.estimator.Estimator(container,
                                    role,
                                    train_instance_count=1,
                                    train_instance_type='ml.c4.xlarge',
                                    output_path=output_location,
                                    sagemaker_session=sess)
```

하이퍼파라미터

```
fm.set_hyperparameters(feature_dim=784,
                       predictor_type='binary_classifier',
                       mini_batch_size=200,
                       num_factors=10)
```

학습 시작

```
fm.fit({'train': s3_train_data})
```

2. TensorFlow 사용하기

Step 1. 학습할 스크립트 준비 (예: tf-train.py)

- TensorFlow 버전: 1.4.1, 1.5.0, 1.6.0, 1.7.0, 1.8.0, 1.9.0, 1.10.0.
- model_fn: 학습할 모델 정의
- train_input_fn: 학습 데이터 전처리 및 로드
- eval_input_fn: 평가 데이터 전처리 및 로드
- serving_input_fn: 예측 모델에 전달할 feature 정의

Step 2. 준비한 스크립트를 `sagemaker.tensorflow.TensorFlow` estimator를 통해 실행

```
from sagemaker.tensorflow import TensorFlow

tf_estimator = TensorFlow('tf-train.py', role='SageMakerRole',
                         training_steps=10000, evaluation_steps=100,
                         train_instance_count=1, train_instance_type='ml.p2.xlarge')
tf_estimator.fit('s3://bucket/path/to/training/data')
```

3. 자동 모델 튜닝 사용하기

Estimator

```
estimator = MXNet(entry_point='mnist.py',
                   role=role,
                   train_instance_count=1,
                   train_instance_type='ml.m4.xlarge',
                   sagemaker_session=sagemaker.Session(),
                   base_job_name='DEMO-hpo-mxnet',
                   hyperparameters={'batch_size': 100})
```

Configuration

```
objective_metric_name = 'Validation-accuracy'
metric_definitions = [{ 'Name': 'Validation-accuracy',
                       'Regex': 'Validation-accuracy=(\d{0}\.\d{1})+' }]
```

```
hyperparameter_ranges = {'optimizer': CategoricalParameter(['sgd', 'Adam']),
                         'learning_rate': ContinuousParameter(0.01, 0.2),
                         'num_epoch': IntegerParameter(10, 50)}
```

Training Jobs

```
tuner = HyperparameterTuner(estimator,
                             objective_metric_name,
                             hyperparameter_ranges,
                             metric_definitions,
                             max_jobs=9,
                             max_parallel_jobs=3)
```

```
tuner.fit({'train': train_data_location, 'test': test_data_location})
```

Resulting Models

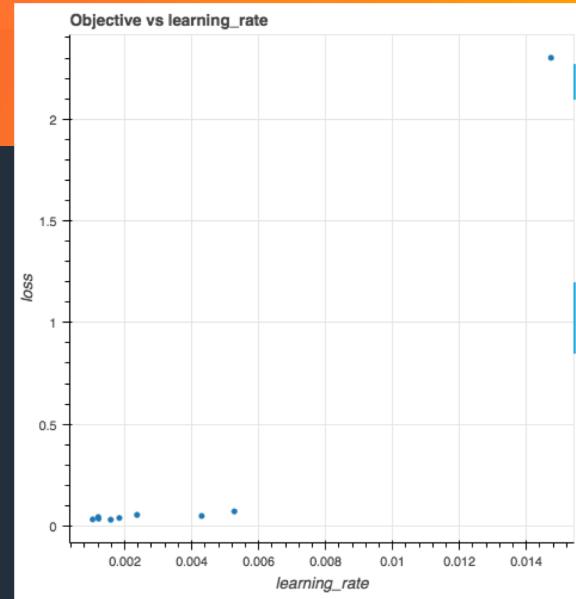
	FinalObjectiveValue	TrainingElapsedTimeSeconds	TrainingEndTime	TrainingJobName	TrainingJobStatus	TrainingStartTime	learning_rate	num_epoch
1	0.9798	136.0	2018-06-08 02:51:31+00:00	180608-0239-008-faeaf8f	Completed	2018-06-08 02:49:15+00:00	0.198522	43.0
0	0.9793	129.0	2018-06-08 02:51:32+00:00	180608-0239-009-eff687d94	Completed	2018-06-08 02:49:23+00:00	0.200000	43.0
2	0.9792	139.0	2018-06-08 02:51:34+00:00	180608-0239-007-b7539e25	Completed	2018-06-08 02:49:15+00:00	0.191755	46.0
4	0.9753	142.0	2018-06-08 02:47:21+00:00	180608-0239-005-ac266f7f	Completed	2018-06-08 02:44:59+00:00	0.052552	46.0
6	0.9719	84.0	2018-06-08 02:42:20+00:00	180608-0239-003-dbcc483c	Completed	2018-06-08 02:40:56+00:00	0.129839	11.0
3	0.9703	88.0	2018-06-08 02:47:16+00:00	180608-0239-006-0a7ab22	Completed	2018-06-08 02:45:48+00:00	0.108355	13.0
5	0.7775	171.0	2018-06-08 02:46:57+00:00	180608-0239-004-a0483ee6	Completed	2018-06-08 02:44:06+00:00	0.067303	37.0
7	0.2052	122.0	2018-06-08 02:43:30+00:00	180608-0239-002-d87c007d	Completed	2018-06-08 02:41:28+00:00	0.172863	15.0
8	0.1010	105.0	2018-06-08 02:43:09+00:00	180608-0239-001-8d691526	Completed	2018-06-08 02:41:24+00:00	0.125656	12.0

4. 자동 모델 튜닝 결과 분석하기

Amazon SageMaker > Hyperparameter tuning jobs

Hyperparameter tuning jobs

Name	Status	Training completed/total	Creation time	Duration
sagemaker-tensorflow-181014-2316	Completed	9 / 9	Oct 14, 2018 23:16 UTC	13 minutes
sagemaker-tensorflow-181013-0658	Completed	9 / 9	Oct 13, 2018 06:58 UTC	14 minutes
sagemaker-tensorflow-181013-0655	Stopped	0 / 3	Oct 13, 2018 06:55 UTC	3 minutes
sagemaker-tensorflow-181013-0631	Completed	9 / 9	Oct 13, 2018 06:31 UTC	13 minutes
sagemaker-tensorflow-181009-0215	Completed	9 / 9	Oct 09, 2018 02:15 UTC	13 minutes
sagemaker-tensorflow-181008-0537	Completed	9 / 9	Oct 08, 2018 05:37 UTC	an hour
xgboost-tuningjob-15-03-44-47	Completed	20 / 20	Aug 15, 2018 03:45 UTC	26 minutes
sagemaker-tensorflow-180815-0327	Completed	9 / 9	Aug 15, 2018 03:27 UTC	22 minutes
sagemaker-tensorflow-180802-0432	Completed	9 / 9	Aug 02, 2018 04:32 UTC	22 minutes
sagemaker-tensorflow-180801-0143	Completed	30 / 30	Aug 01, 2018 01:43 UTC	an hour
sagemaker-tensorflow-180801-0143	Completed	30 / 30	Aug 01, 2018 01:43 UTC	an hour
sagemaker-tensorflow-180801-0143	Completed	30 / 30	Aug 01, 2018 01:43 UTC	an hour
sagemaker-tensorflow-180801-0143	Completed	30 / 30	Aug 01, 2018 01:43 UTC	an hour
sagemaker-tensorflow-180801-0143	Completed	30 / 30	Aug 01, 2018 01:43 UTC	an hour
sagemaker-tensorflow-180801-0143	Completed	30 / 30	Aug 01, 2018 01:43 UTC	an hour



핸즈온 주의사항

1. AWS 계정이 반드시 필요합니다.
2. Browser는 최신 버전의 Chrome, Firefox를 권장합니다.
3. 실습 리소스 생성시 Seoul 리전을 사용합니다.
4. AWS 콘솔 언어는 화면 좌측 하단에서 변경 가능합니다.
5. 실습 종료 후 반드시 리소스를 삭제하시기 바랍니다.



DEV DAY

Q&A

- 세션 후, 설문에 참여해 주시면 행사 후 소정의 선물을 드립니다.
- #AWSDevDay 해시 태그로 의견을 남겨주세요!