

Self-Study-On-SageMaker

마지막 업데이트: 2021. 01. 19

작성자: 문곤수

이 문서의 목적은 혼자 SageMaker 의 학습을 하기 위한 자료 및 링크가 있습니다. 아래와 같은 큰 목차에 관련 정보가 있으니 참고 하세요.

- SageMaker 선수 지식 및 101
- SageMaker 입문
- SageMaker 기본
- 유용한 자료 (백서 등)
- 체계적인 교육
- 대학교 AWS SageMaker 커리큘럼 수업 예시

SageMaker 선수 지식 및 101

- AWS Cloud 일반 지식 (S3, EC2, IAM 등) (초급/중급 정도 수준)
- Python 코딩 (Pandas, Numpy 패키지 초급/중급 정도 수준)
- SageMaker Video (1분 40초)
 - 전체적인 간략 소개 입니다.
 - <https://www.youtube.com/watch?v=jSrg1BgQzRo>
- AWS SageMaker 공식 웹 사이트
 - 이것은 자주 봐야할 곳 입니다. 처음에는 간단히 보시기 바랍니다.
 - 전반적인 소개 및 가능한 다양한 리소스 정보를 접할 수 있습니다.
 - <https://aws.amazon.com/ko/sagemaker/>
- 주피터 노트북 1회 실행 (아래에서 1개만 수행하시면 됩니다.)
 - SageMaker 10분 튜토리얼
 - 세이지 메이커 콘솔에서 노트북 인스턴스를 설치하고, 제공된 파이썬 코드를 Copy & Paste를 하면 모델 빌딩, 구축, 배포를 확인할 수 있습니다. 코드의 이해는 일단 지나치시고, 전체적으로 동작하는 원리만 아는 것으로 초점을 맞추어 주세요.
 - 아래를 진행할 시에 리전은 us-west-2, us-east-1, us-east-2, eu-west-1 중에 하나를 선택해야 합니다. : 빌트인 내장 알고리즘 컨테이너를 위 리전에서 가져옴
 - 자주 발생하는 에러는 버킷 생성이 안되어 납니다. 참고 하세요.
 - <https://aws.amazon.com/ko/getting-started/hands-on/build-train-deploy-machine-learning-model-sagemaker/>
 - 다른 노트북 인스턴스 사용 예제
 - Amazon SageMaker 노트북 인스턴스 사용
 - https://docs.aws.amazon.com/ko_kr/sagemaker/latest/dg/nbi.html
 - <https://www.ciokorea.com/tags/11396/R/118326>
 - <https://scikitlearn.tistory.com/73>
 -
- 머신 러닝 기본 (기타 자료를 보셔도 됩니다.)
 - 생활 코딩: 머신러닝

- <https://opentutorials.org/course/4548>
- 다커 컨테이너 동작 방식 - 이것은 옵션입니다. 지나 치셔도 됩니다.
 - Hello Docker
 - <https://github.com/mullue/hello-docker>
 -

SageMaker 입문

목표: 세이지 메이커의 기본 구조와 사용법 알기

- SageMaker 소개 유튜브 비디오 (한글)
 - SageMaker Overview (소개 동영상) - <https://youtu.be/jF2BN98KBlg>
 - SageMaker demo - <https://youtu.be/miIVGlq6OUk> (1시간 데모에 많은 내용을 압축해서 다루고 있습니다. 반복해서 보시거나 돌려보기로 차근차근 보셔도 괜찮습니다.)
- Introduction to Amazon SageMaker (12분)
 - 코세라 공식 세이지 메이커 소개 비디오
 - <https://www.coursera.org/lecture/aws-machine-learning/introduction-to-amazon-sagemaker-QugTh>
- Amazon SageMaker Deep Dive Series (아래 두개만 일단 보셔도 됩니다. 각 비디오가 약 10-20분 사이 입니다.)
 - https://www.youtube.com/playlist?list=PLhr1KZpdzucOr_6j_zmSrvYnLUtgqsZz
 - Fully-Managed Notebook Instances with Amazon SageMaker - a Deep Dive
 - Built-in Machine Learning Algorithms with Amazon SageMaker - a Deep Dive
- SageMaker 최초 실습
 - 한글 워크샵 사이트(<https://www.sagemaker-workshop-kr.com/kr>)에서 다음 두 모듈을 진행합니다.(빌트인 알고리즘 활용)
 - ◦ 모듈1 SageMaker > S3 bucket과 노트북 생성하기 - https://www.sagemaker-workshop-kr.com/kr/sagemaker/_module_1.html
 - ◦ 모듈2 Linear Learner MNIST - https://www.sagemaker-workshop-kr.com/kr/sagemaker/_module_2.html
 - 기본 Tabular 데이터를 사용: Tabular 데이터를 모델 빌딩, 훈련, 배포, 추론을 해볼 수 있는 예제 입니다.
 - 다이렉트 마케팅 (SageMaker 내장 알고리즘 XGBoost 사용)
 - <https://github.com/mullue/xgboost> (1번 노트북 진행)

SageMaker 기본

목표: 세이지 메이커의 유스 케이스별 사용 알기

- 공식 예제 사이트: Amazon SageMaker Example Notebooks (영어)
 - 공식 SageMaker 의 공식 예제 설명 문서 입니다. 모든 공식 샘플이 있기에 필요할때 마다 필요한 것을 찾아서 보시면 되겠습니다.
 - 단계 별, 주제 별로 모든 공식 예제가 정리 되어 있습니다. 1. Getting Started, 2. SageMaker Studio, 3. AutoPilot 4.Ingest Data, 5. Label Data, 6.Prep Data, 7.Feature Store 8.Training 9.Inference 10.Frameworks 11.

Workflows 12. Advanced Examples 13. Community Examples 의 대 카테고리 밑에 샘플 예제의 설명 및 코드가 있습니다.

- <https://sagemaker-examples.readthedocs.io/en/latest/index.html>
 - 공식 Git: <https://github.com/aws/amazon-sagemaker-examples>
- 공식 ML 블로그: 적용 사례, 다른 aws서비스와의 통합관련 예제나 기술 팁등 다양한 주제들이 다루어집니다.
 - aws ML blog(<https://aws.amazon.com/ko/blogs/machine-learning/>)
- Use Case 별 예제 및 자료
 - 한글 예제 사이트
 - 데이터 분석과 머신러닝
 - 클라우드 기반 데이터 분석 및 인공 지능을 위한 비즈니스 혁신: <https://www.youtube.com/watch?v=24YgdrJ9r-A>
 - 수백만 사용자 대상 기계 학습 서비스를 위한 확장 비법 - <https://www.youtube.com/watch?v=RYzvzuOCU>
 - 테블러 (테이블 데이터)
 - 배송 시간 예측하기: 테이블(Tabular) 데이터 및 AutoGluon, XGBoost 알고리즘을 이용 (AutoGluon 사용에 초점)
 - <https://github.com/gonsoomoon-ml/predict-delivery-time>
 - SageMaker로 싱글 타임시리즈 시계열 예측 워크샵: <https://github.com/gonsoomoon-ml/linear-regression-forecast>
 - Inference Pipeline을 이용한 고객 이탈 예측 모델 및 평가 (Churn Prediction Model): <https://github.com/gonsoomoon-ml/churn-prediction-workshop>
 - 자연어 (NLP)
 - 토픽 모델링을 사용한 온라인 상품 부정 리뷰 분석: <https://github.com/gonsoomoon-ml/topic-modeling>
 - 빌트인 알고리즘을 이용한 한글처리 - <https://github.com/daekeun-ml/blazingtext-workshop-korean>
 - BERT 이용한 한글처리 - <https://github.com/daekeun-ml/kobert-workshop>
 - 영상 (Vision)
 - 자동차 번호판 인식 문제 해결에 SageMaker 적용해보기 - <https://github.com/mullue/lab-custom-model-anpr>
 - 기타
 - Tensorflow 활용실습 (Tensorflow 2.0 script mode와 stepfunctions사용하기)
 - <https://github.com/mullue/sm-tf2>
 - Kubernetes와 Sagemaker를 활용하여 기계학습 워크로드 관리하기: <https://www.youtube.com/watch?v=6sogVHw9jZ4>
 - SageMaker와 EFS 연결하기 - <https://aws.amazon.com/ko/blogs/machine-learning/speed-up-training-on-amazon-sagemaker-using-amazon-efs-or-amazon-fsx-for-lustre-file-systems/>
 - Kubernetes Kubeflow와의 통합 - <https://aws.amazon.com/ko/blogs/machine-learning/introducing-amazon-sagemaker-components-for-kubeflow-pipelines/>
 - Elasticsearch와 연계한 이미지 검색 시스템 - <https://aws.amazon.com/ko/blogs/machine-learning/building-a-visual-search-application-with-amazon-sagemaker-and-amazon-es/>
 - Ground Truth를 이용한 3D 레이블링 - <https://aws.amazon.com/ko/blogs/machine-learning/labeling-data-for-3d-object-tracking-and-sensor-fusion-in-amazon-sagemaker-ground-truth/>

- DLAMI를 이용한 분산학습 - <https://aws.amazon.com/ko/blogs/machine-learning/multi-gpu-distributed-deep-learning-training-at-scale-on-aws-with-ubuntu18-dlami-efa-on-p3dn-instances-and-amazon-fsx-for-lustre/>
 - 유스케이스별로 deploy하여 확인할 수 있는 케이스는 aws solutions ML 섹션을 참고하십시오.
 - - <https://aws.amazon.com/solutions/implementations/?solutions-all.sort-by=item.additionalFields.sortDate&solutions-all.sort-order=desc&awsf.AWS-Product%20Category=tech-category%23ai-ml>
- 워크샵
 - 동일 사이트의 reference 페이지 (<https://www.sagemaker-workshop-kr.com/kr/references.html>)

유용한 자료 (백서 등)

- ML 프로젝트 전반에 대한 이해가 필요하시면 다음 AWS ML백서들이 도움이 되실 수 있습니다.
 - AWS 딥러닝 가이드 백서입니다.
 - https://d1.awsstatic.com/whitepapers/Deep_Learning_on_AWS.pdf?did=wp_card&trk=wp_card
 - ML Project 수행에 대한 베스트 프랙티스 가이드입니다.
 - -https://d1.awsstatic.com/whitepapers/aws-managing-ml-projects.pdf?did=wp_card&trk=wp_card
 - AWS ML Well Architect Framework (WAF)
 - ML의 WAF 내용 백서입니다.
 - <https://d1.awsstatic.com/whitepapers/architecture/wellarchitected-Machine-Learning-Lens.pdf>
- 외부 사이트
 - medium.com 에서 검색해서 사용
 - 비전 딥러닝
 - 라온 피블 블로그
 - <https://m.blog.naver.com/PostList.nhn?blogId=laonple>
 - NLP 딥러닝
 - 딥러닝을 이용한 자연어 처리 입문
 - <https://wikidocs.net/book/2155>

외부 체계적인 교육

- Coursera Sagemaker - <https://www.coursera.org/lecture/aws-machine-learning/introduction-to-amazon-sagemaker-QugTh>
- Coursera computer vision - <https://www.coursera.org/learn/aws-computer-vision-gluoncv>
- Deep Learning Specialisation (코세라)
 - 총 5개의 코스로 구성됨: Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization, Structuring Machine Learning Projects, Convolutional Neural Networks(CNN), Sequence Models(RNN)
 - <https://www.coursera.org/specializations/deep-learning>
- - Udacity nanodegree 코스 - <https://www.udacity.com/course/machine-learning-engineer-nanodegree--nd009t>

대학교 AWS SageMaker 커리큘럼 수업 예시

1. 전체 과정 소개

2. AWS 활용 준비하기, AWS Educate 가입 등

3 실습 환경 구축 SageMaker 소개, Jupyter Notebook 인스턴스 생성 실습

- 기초 핸즈온 수행 (위의 SageMaker 선수 지식 및 101 과정 참고)

4 기계학습 소개 이론

5 선형회귀분석 이론 및 Jupyter Notebook 코드리뷰 이론 + 실습

- Regression with Amazon SageMaker Linear Learner algorithm
 - https://github.com/aws/amazon-sagemaker-examples/blob/master/introduction_to_amazon_algorithms/linear_learner_abalone/Linear_Learner_Regr_ession_csv_format.ipynb

6 트리 이론 및 Jupyter Notebook 코드리뷰 이론 + 실습

- *Gradient Boosted Trees를 이용하는 지도학습: 편향된 클래스의 이진 분류 예측문제 해결*
- https://github.com/mullue/xgboost/blob/master/1.xgboost_direct_marketing_sagemaker.ipynb

7 인공신경망 이론 및 Jupyter Notebook 코드리뷰 이론 + 실습

8 k-means 이론 및 Jupyter Notebook 코드리뷰 이론 + 실습

- Analyze US census data for population segmentation using Amazon SageMaker
 - https://github.com/aws/amazon-sagemaker-examples/blob/master/introduction_to_applying_machine_learning/US-census_population_segmentation_PCA_Kmeans/sagemaker-countycensusclustering.ipynb

9 텍스트 데이터 분석 이론 및 Jupyter Notebook 코드리뷰 이론 + 실습

- 토픽 모델링을 사용한 온라인 상품 부정 리뷰 분석
 - <https://github.com/gonsoomoon-ml/topic-modeling>

10 Amazon SageMaker 를 활용한 모델 배포 실습

- TensorFlow 2 프로젝트 워크플로우를 SageMaker에서 실행하기
 - <https://github.com/mullue/sm-tf2/blob/master/tf-2-workflow.ipynb>