Cloud computing

Project: Linear regression model for power prediction using AWS Sagemaker

The goal of this project was to create a model capable of predicting power generated at a Combined Cycle Power Plant (CCPP), using dataset of 9569 measurement points of ambient parameters. The data consisted of: Atmospheric temperature and pressure, relative humidity and exhaust vacuum. Data was taken from:

<https://archive.ics.uci.edu/ml/datasets/combined+cycle+power+plant>

Data was downloaded in a form of a csv file.

Tools used in this project were:

* Amazon Sagemaker – AWS platform for machine learning
* Amazon Simple Storage Service (S3)

Sagemaker is a perfect tool to use in this case. It allows user to work in a Jupyter notebook on a chosen instance without having to download or configure it by hand. Additionally there are multiple algorithms available to the user which simplifies the process of creating a model. The algorithm used in this project was Linear learner.

The code was written in Python 3. The steps in a process were:

1. loading the dataset
2. splitting it in 3 to training:

* training which was used to train a model
* validation to test the model during the process of training and adjust the model
* test to compare model predictions with the remaining data

1. uploading files to created S3 bucket
2. loading the linear learner algorithm
3. training the model at a ml.c4.xlarge type instance
4. creating the endpoint of the trained model
5. testing the model
6. deleting the endpoint to release the instance

The accuracy of the model was 99.92 %.