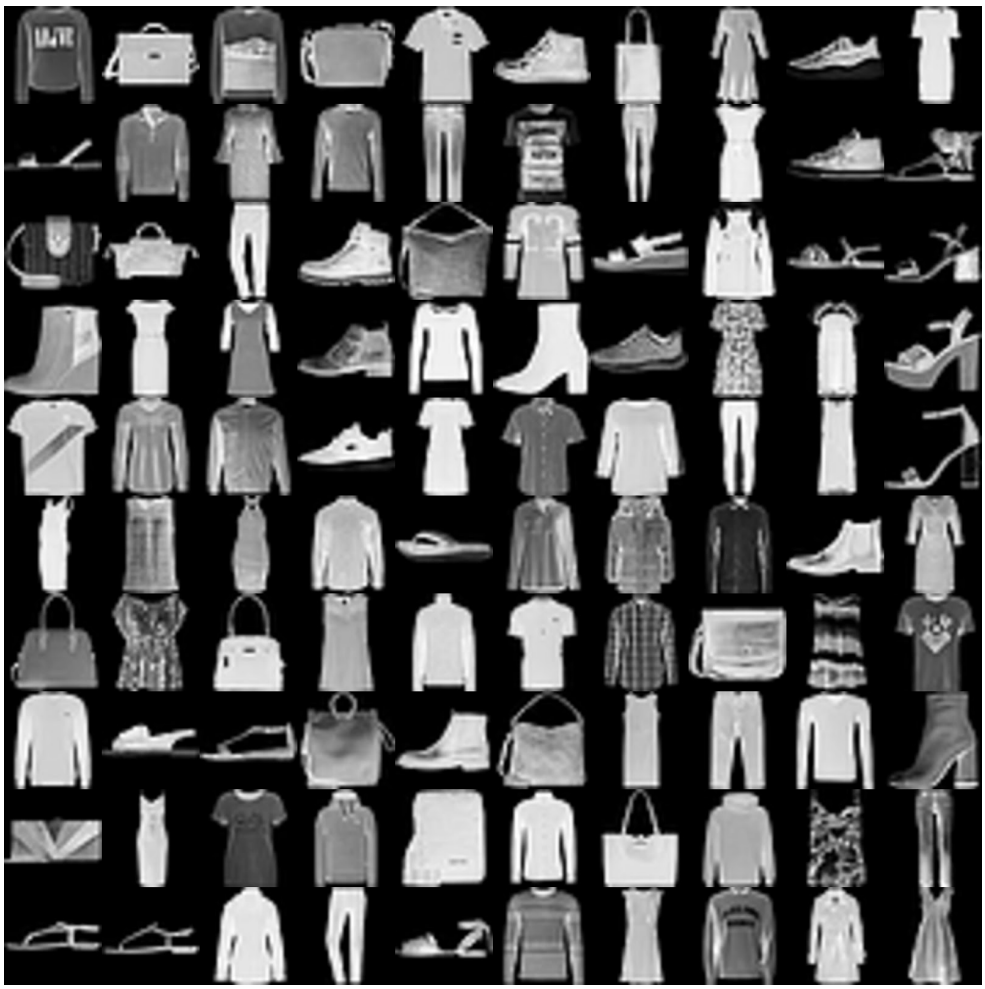

COMP47590

Advanced Machine Learning

Lab Task

Introduction

Zalando, a fashion and technology company with a large presence in Dublin, have released MNIST Fashion, an alternative to the famous MNIST hand written digit classification dataset. The dataset contains 70,000 28x28 pixel grayscale images of fashion items of 10 different kinds: (0) T-shirt/top, (1) Trouser, (2) Pullover, (3) Dress, (4) Coat, (5) Sandal, (6) Shirt, (7) Sneaker, (8) Bag, and 9) Ankle boot. The task associated with this is to build a classification model that can recognise these different fashion items.

Tasks

Perform the following tasks:

scikit-learn在anaconda自带

1. Download the MNIST fashion training dataset from Kaggle at https://www.kaggle.com/zalando-research/fashionmnist
2. Create a Jupyter python notebook in which to load data and build and evaluate machine leanring models using scikit-learn
3. Load and explore the MNIST fashion training dataset
4. Dived the MNIST fashion training dataset into a train and validation datasets (using 70:30 split)
5. Build a decision tree model to perform the MNIST fashion classification task
6. Evaluate the performance of this tree on the MNIST fashion validation set created previously
7. User a grid search to tune the hyperparaemters for this model
8. Experiment with different machine learning algorithms to see if you can build a better model (try ensembles, bagging, boosting, random forests, ...)
9. Experiment with feature engineering that you could perform in order to do better at this problem.
10. Download the MNIST fashion test dataset from Kaggle at https://www.kaggle.com/zalando-research/fashionmnist
11. Evaluate your "best" model on the test dataset