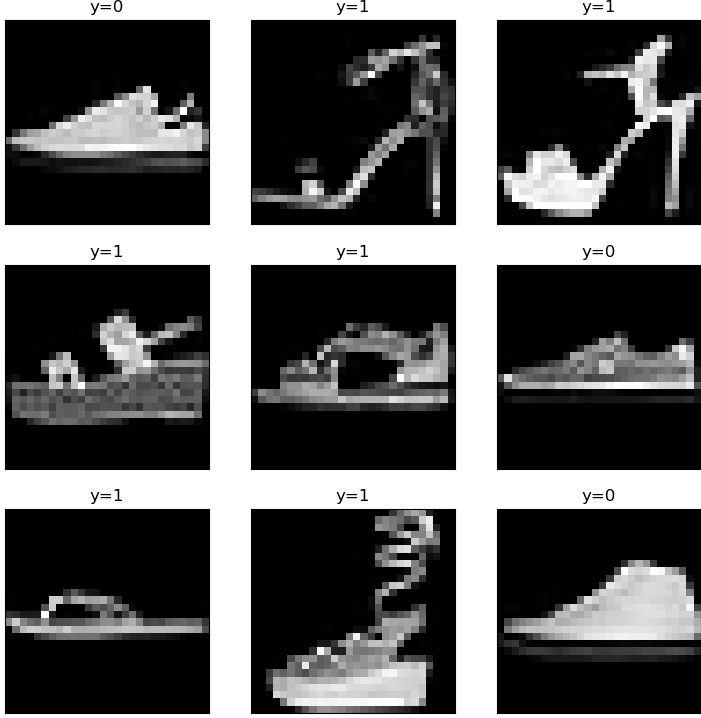
Exploring ML methods for Image Classification

**Description:**

Today, we're going to explore multiple machine learning methods that have been used in the field of computer vision to classify images. Although the state of the art is much more advanced and relies on deep learning ([check this link](https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/)), to keep things understandable, we’re sticking with Logistic Regression, Decision Trees, and Multi Layer Perceptrons (neural networks).

In this lab, we’re looking at the sneaker vs sandal dataset:



Each example is a 28x28 image, which is reshaped into a 784-dimensional vector.

These are from the larger [Fashion MNIST dataset](https://github.com/zalandoresearch/fashion-mnist/), made public originally by Zalando Research.

**Support Code:**

We’ve set up reading in the data. Your job is to use the official documentation of sklearn’s model classes to implement three different classification models on a toy dataset and report your achieved accuracy for each model. If you have time, for extra credit, you may try to improve results by augmenting your data/features, messing with the configurations of your models, or trying an entirely new architecture available in sklearn’s documentation.

**Your Task:**

Your goal is to make three different changes to the existing system to aid in choosing between the models. Check the helpful resources section for documentation on the three classifiers you are supposed to implement.

1. Train a logistic regression model on the train set and predict on the test set using the trained model
2. Train a decision tree classifier model on the train set and predict on the test set using the trained model
3. Train a multi layer perceptron model on the train set and predict on the test set using the trained model
4. Record all of the results and observe what model performs the best
5. **Extra Credit:** If you have extra time, or want to learn more about machine learning, do whatever you want to try to improve the accuracy of your best model! Use google- Data augmentation and feature engineering are good topics to look up. Include the code and report your results.

**Files Given:**

x\_train.csv – training set containing 9,998 examples of either a sneaker or a sandal

y\_train.csv – true labels for the train set. 0 is sneaker, 1 is sandal

x\_test.csv – test set containing 2002 examples of either a sneaker or a sandal

y\_test.csv – true labels for the test set. 0 is sneaker, 1 is sandal

**Helpful Resources:**

[Log Loss Documentation](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.log_loss.html#sklearn.metrics.log_loss)

[Logistic Regression Documentation](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

[Decision Tree Classifier Documentation](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier)

[MLP Classifier Documentation](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html)

[general sklearn api reference](https://scikit-learn.org/stable/modules/classes.html)

**Submit your results on replit!**