**Description**

**Start here if...**

You have some experience with R or Python and machine learning basics, but you’re new to computer vision. This competition is the perfect introduction to techniques like neural networks using a classic dataset including pre-extracted features.

**Competition Description**

MNIST ("Modified National Institute of Standards and Technology") is the de facto “hello world” dataset of computer vision. Since its release in 1999, this classic dataset of handwritten images has served as the basis for benchmarking classification algorithms. As new machine learning techniques emerge, MNIST remains a reliable resource for researchers and learners alike.

In this competition, your goal is to correctly identify digits from a dataset of tens of thousands of handwritten images. We’ve curated a set of tutorial-style kernels which cover everything from regression to neural networks. We encourage you to experiment with different algorithms to learn first-hand what works well and how techniques compare.

**Practice Skills**

* Computer vision fundamentals including simple neural networks
* Classification methods such as SVM and K-nearest neighbors

**Acknowledgements**

More details about the dataset, including algorithms that have been tried on it and their levels of success, can be found at <http://yann.lecun.com/exdb/mnist/index.html>. The dataset is made available under a [Creative Commons Attribution-Share Alike 3.0 license](https://creativecommons.org/licenses/by-sa/3.0/).

**Tutorial**

**Kaggle Learn**

[**Kaggle Learn**](https://www.kaggle.com/learn) offers hands-on courses for most data science topics. These short courses prepare you with everything you need to start your own projects, including deep learning and computer vision projects.

The [**Deep Learning Course**](https://www.kaggle.com/learn/deep-learning) will give you everything you need to succeed in this competition and others like it.

**Other Python Tutorials**

[**Deep neural network the Keras way**](https://www.kaggle.com/poonaml/digit-recognizer/deep-neural-network-keras-way)

* Covers pre-processing including feature standardization and one-hot encoding
* Implements an artificial neural network approach using Keras

[**Simple deep MLP with Keras**](https://www.kaggle.com/fchollet/digit-recognizer/simple-deep-mlp-with-keras/code)

* A straightforward implementation of MLP (multi-layer perceptron) in Keras
* Learn Keras from the author himself, Francois Chollet!

[**An introduction to dimensionality reduction**](https://www.kaggle.com/arthurtok/digit-recognizer/interactive-intro-to-dimensionality-reduction)

* Introduces and compares PCA, LDA, and t-SNE dimensionality reduction techniques
* Uses the Plotly library for intuitive, interactive visualizations

**Other R Tutorials**

[**Random forest benchmark**](https://www.kaggle.com/benhamner/digit-recognizer/random-forest-benchmark/code)

* A minimal example implementing the random forest algorithm

[**Build your own neural network in R**](https://www.kaggle.com/russwill/digit-recognizer/build-your-own-neural-network-in-r)

* Implements a simple 2-layer neural network from scratch
* Based on the CS231n course offered by Stanford

[**Minimum distance classifier**](https://www.kaggle.com/olhacher/digit-recognizer/minimum-distance-classifier)

* Uses minimum distance as a simple approach to classification

##### Evaluation

### Goal

The goal in this competition is to take an image of a handwritten single digit, and determine what that digit is.  
For every in the test set, you should predict the correct label.

### Metric

This competition is evaluated on the categorization accuracy of your predictions (the percentage of images you get correct).

### Submission File Format

The file should contain a header and have the following format:

ImageId,Label  
1,0  
2,0  
3,0  
etc.

##### Frequently Asked Questions

### What is a Getting Started competition?

Getting Started competitions were created by Kaggle data scientists for people who have little to no machine learning background. They are a great place to begin if you are new to data science or just finished a MOOC and want to get involved in Kaggle.

Getting Started competitions are a non-competitive way to get familiar with Kaggle’s platform, learn basic machine learning concepts, and start meeting people in the community. They have no cash prize and are on a rolling timeline.

### How do I create and manage a team?

When you accept the competition rules, a team will be created for you. You can invite others to your team, accept a merger with another team, and update basic information like team name by going to the [Team page](https://www.kaggle.com/c/digit-recognizer/team).

We've heard from many Kagglers that teaming up is the best way to learn new skills AND have fun. If you don't have a teammate already, consider asking if anyone wants to team up in the [discussion forum](https://www.kaggle.com/c/digit-recognizer/discussion).

### What are Notebooks?

Kaggle Notebooks is a cloud computational environment that enables reproducible and collaborative analysis. Kernels supports scripts in R and Python, Jupyter Notebooks, and RMarkdown reports. Go to the [Notebooks tab](https://www.kaggle.com/c/digit-recognizer/notebooks) to view all of the publicly shared code on this competition. For more on how to use Notebooks to learn data science, visit Kaggle's [Learn Courses](https://www.kaggle.com/learn/overview).

### What’s the difference between a private and public leaderboard?

In this competition, because it is a Getting Started competition, there is no difference. We're scoring the entire test set on the Public Leaderboard. And we will refresh the competition every three months, so the Private Leaderboard is irrelevant.

For non-Getting Started Kaggle competitions, there is the concept of a public and private leaderboard to prevent participants from “overfitting” to the leaderboard. If your model is “overfit” to a dataset then it is not generalizable outside of the dataset you trained it on. This means that your model would have low accuracy on another sample of data taken from a similar dataset.

### Why are there perfect scores on the leaderboard?

This competition's test set labels are completely public. So it's likely that some participants will submit perfect submissions. Since there are no prizes, medals, or points associated with your leaderboard ranking in this competition, these scores have little consequence. The rankings are purely for the benefit of users to learn and see how their approach is improving.

"Then can you just clear them out? They're distracting." It's a heavily time-intensive process to manage removals from a continuously live leaderboard. The threshold for what should be removed also becomes arbitrary -- if you remove 1.00 scores, then .99 scores will quickly appear. Since there is no prize consequence, we have chosen to leave the scores as an open sandbox.