Introduction to Machine Learning
Part II
Mehran Behjati
• Recap
• Neural Network
• Tools:
  • Colab
  • pandas
  • Numpy
  • Scikit-learn
  • TensorFlow
  • Edge Impulse
Recap

Importance of data
- Training set
- Validation set
- Test set

Issues:
- Bias
- Unbalanced
- Small

Data cleaning
- Removing unwanted
- Managing outliers
- Handling missing data

Feature selection

Feature extraction

Split dataset

Model selection
- Regression
- Classification
- Clustering

Overfitting
Underfitting

Evaluation matrices
Confusion matrix
Neural Network

• What is Artificial Neural?
• It was inspired by the understanding around how biological neurons work and operate in the human brain.
Deep Neural Network

- Neural Network
- **Deep/dense NN**: NN models with hidden layers, where every layer densely connected with the previous and the next layer
Handwritten digit recognition
### Activation Function

<table>
<thead>
<tr>
<th>Sigmoid</th>
<th>RELU</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(z) = \frac{1}{1 + e^{-z}}$</td>
<td>$g(z) = \max(0, z)$</td>
</tr>
</tbody>
</table>

![Graphs of Sigmoid and RELU activation functions](image)
Feedforward
Cost/Loss/Error Function
Backpropagation

Forwardpass

Backwardpass

\[
\frac{dL}{dx} = \frac{dL}{dz} \frac{dz}{dx}
\]

\[
\frac{dL}{dy} = \frac{dL}{dz} \frac{dz}{dy}
\]
Gradient decent
Epoch
Some of Essential Terminologies

- Neuron/node
- Activation
- Weight
- Bias
- Activation function
- Input, hidden, and output layer
- Forward propagation / Feedforward
- Backpropagation
- Cost/Loss function
- Iteration
- Epoch
How to create a ML project
• An executable document lets you write and run your code on browser
• Already pre-installed
• Connects your notebook to a cloud-based runtime
  • We can use the power of cloud to train our models efficiently
• Easy to share code within Google Drive or GitHub
• Edit collaboratively
• Hardware acceleration, GPUs & TPUs
• The volume of data is rapidly growing
• The importance of processing this data is much more than decades ago
• Excel is good but we need more flexible, powerful and advanced tools
• It is an open-source library that has been developed in Python
• Almost everyone who works on the field of data science needs to know pandas
  • read data from csv
  • work with series
  • arithmetic and statistical operations
  • sort
  • work with data frames (combinations of series)
  • choosing rows and columns
  • filtering
  • optimizing memory
• A famous library among the data scientists for scientific computing in Python.

• Something similar to MATLAB, for array and matrix operations

• Working on numbers, array, and matrices
  • Array creation
  • Indexing, slicing, and iterating
  • Shape manipulation
  • Arrays stacking
  • Other operations, including mathematical, logical, sorting, selecting, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.
• An open source ML library that supports supervised and unsupervised learning
• Provides various tools for model fitting, data preprocessing, model selection, model evaluation, and many other utilities
• Provides dozens of built-in ML algorithms and models
• An open source end-to-end platform for ML
• Provides a comprehensive ecosystem of tools to built ML applications
• Developed and supported by Google to process and analyze data
• Supports wide range of ML algorithms and models
• Provides APIs for different languages, Python, C++, Javascript, ...
• TensorFlow is originally designed for big computing systems
• Is it suitable for TinyML implementation?
Installations

• You can install Python and other tools separately on your device
• Alternatively you can install Anaconda
• A great tool to create ML projects
• Makes model training much easier
• Gives a graphical interface & representation