Fundamentals of Machine Learning

Antonio Jones

13 February 2020

Overview

Model Evaluation

Data Processing & Feature Engineering

Over (& Under)-fitting

Machine Learning Workflows

Evaluating Machine Learning Models

Branches of Machine Learning:

A. Unsupervised Learning	B. Unsupervised Learning
C. Reinforcement Learning	D. Reinforcement Learning

Evaluating Machine Learning Models

Data Segmentation Strategies:

A. Simple Hold-out Validation
B. K-Fold Validation
C. Iterated K-Fold Validation with Shuffling

Preparing the Input Data

Data Preprocessing for Neural Networks

i. Vectorization

ii. Value Normalization

iii. Missing Values

~N(0,1)

Tensors

[...,x18, x19, ?? , x21, x22,...]

Feature Engineering

Outside Knowledge

Applied to the ML algorithms

More elegant solutions that require less data

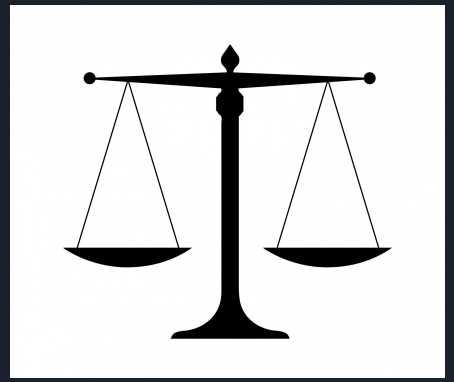
Over- & Underfitting the Model

Over-fitted Models

Learn misleading or irrelevant patterns for training data

Perform well on training data but unable to generalize and predict with test data

Generalization



Optimization

Under-fitted Models

Unable to model training data; relevant patterns in the data not evaluated

Training accuracy (or other evaluating metric) is low while loss is high

Approaches to resist overfitting:

i. Reduce the network's capacity

ii. Add weight regularization

iii. Add dropout

Machine Learning Workflows

1. Define the problem
2. Choose the measure of success
3. Decide on an evaluation protocol
4. Prepare your data
5. Develop a model that does better than a baseline
6. Scale up: develop a model that overfits
7. Regularizing your model and tuning your hyperparameters