Machine learning basics

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — 861fb9d0

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- Is it currently possible to batch process in parallel? If so, how is the network updated?

1 Overview

1.1 Math Basics

- Matrices
- Vectors
- Operations (addition, subtraction, multiplication by scalar, matrix/vector and matrix/matrix multiplication, inverse, transpose)

1.2 Core Concepts

- Cost function
- Gradient descent
- Normal equation
- Classification (binary vs multiclass)
    - One vs all/rest
- Linear/non-linear boundaries
- Under/over-fitting
    - Reduce the number of features
    - Regularization

1.3 Neural Networks

- Layer
- Bias
- Weight
- Input/output
- Activation
- Forward/backward propagation

1.4 Data Sets

- Training/test/validation sets

1.5 Diagnosis of Learning Algorithm

- Bias (underfit) vs variance (overfit)
- Learning curve (error based on training set size)

1.6 Fixes

- Get more training examples → fix high variance
- Try smaller sets of features → fix high variance
- Try getting additional features → fix high bias
- Try adding polynomial features → fix high bias
- Try decreasing lambda → fix high bias
- Try increasing lambda → fix high variance

neural networks architecture

- small neural network -> prone to underfitting
- large neural network -> prone to overfitting

how to prioritize your work

- start with a simple algorithm that can be implemented quickly
    - implement it and test it on cross-validation data
- plot learning curves to decide if more data, features, etc. are likely to help
- do error analysis: manually examine the examples (in the cross validation set) that your algorithm made errors on. See if you spot any systematic trend in what type of examples it is making errors on
- use pareto rule (80/20)
- use numerical evaluation to determine if a change is an improvement or not

skewed classes

- precision/recall
- precision = true positive/# predicted positive = (true+/(true+ + false+))
- recall = true positive/# actual positive = (true +/(true+ + false-))
- f1 score = $2pr/(p+r)$

support vector machine
kernels

- linear
- gaussian

unsupervised learning

- clustering
    - k-means (cluster assignment, move centroid to the cluster's mean location)
    - elbow method (choose the number of clusters automatically)
- dimensionality reduction (data compression)
    - principal component analysis

To avoid overfitting, the number of parameters estimated from the data must be considerably less than the number of data points

- Linear regression: estimate the coordinates of a value
- Logistic regression: answer a yes/no question
- Softmax classification: answer a multiple choice question

2 Procedure

- Define the task

- Obtain data (write a data fetcher)
- Prepare data
- Implement network architecture
  - Define cost, error and optimize computation methods
- Feed data into network

3 Drawing a rectangle over an image

4 See also

5 References

- http://www.erogol.com/machine-learning-pathway/