Mathematics for Machine Learning

Dina Machuve (PhD)

IndabaX Tanzania 2019, UDOM-CIVE

11th April 2019

Reference & Acknowledgment

- Deisenroth et al.: Mathematics for Machine Learning, Chapter 5
  https://mml-book.com
- Deisentroth, M. (2018); Foundations of Machine Learning 2018/2019(AMMI) Course, AIMS Rwanda.

Outline

- Introduction
- Vector Calculus
  - Scalar differentiation
  - Partial derivatives
  - Jacobian
  - Chain rule
  - Derivatives of matrices w.r.t. matrices
  - Gradients in a multi-layer neural network

Applications of Machine Learning

- Product Recommendations
- Self driving cars
- Virtual Personal Assistants
- Diseases detection
- Predictions while Commuting
- Videos Surveillance
- Social Media Services
- Email spam and malware filtering
- Online customer support
- Search Engine result refining

Mathematical Concepts in Machine Learning

- Linear Algebra
- Analytic Geometry
- **Vector Calculus**
- Statistics and Probability Theory
- Optimization

Vector Calculus Overview

Algorithms in Machine Learning:

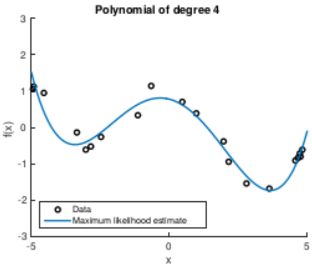
- optimizing an objective function with respect to a set of desired model parameters $\rightarrow$ Optimization problem
- Examples
  - linear regression $\rightarrow$ optimize linear weight parameters to maximize the likelihood
  - neural network auto-encoders for dimensionality reduction and data compression $\rightarrow$ parameters are the weights and biases of each layer, we minimize a reconstruction error by repeated application of the chain-rule
  - Gaussian mixture models for modeling data distributions, $\rightarrow$optimize the location and shape parameters of each mixture component to maximize the likelihood of the model

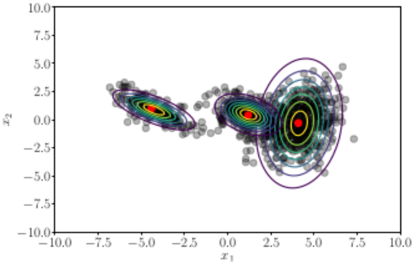
Vector calculus focuses on how to compute gradients of functions:

a. regression (curve fitting)

b. density estimation i.e. modeling data distributions

Role of Vector Calculus

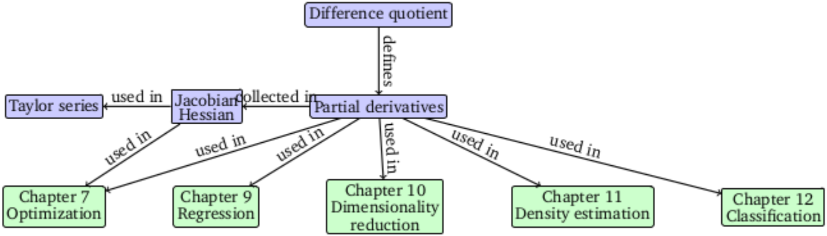




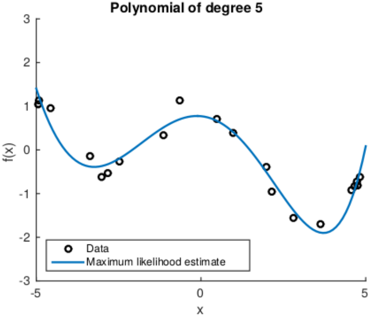
(a)Regression problem: Find parameters, such that the curve explains the observations(circles) well.

(b)Density estimation with a Gaussian mixture model: Find means and covariances, such that the data (dots) can be explained well.

Vector Calculus: A Mind Map



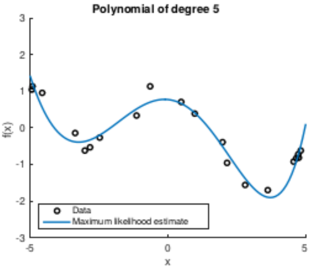
Curve Fitting (Regression) in Machine Learning (1)



- Setting: Given inputs x, predict outputs/targets y
- Model f that depends on parameters θ. Examples:
  - Linear model: $f(x, \theta) = \theta^T x, \quad x, \theta \in \mathbb{R}^D$
  - Neural network: $f(x, \theta) = NN(x, \theta)$

Curve Fitting (Regression) in Machine Learning (2)

- Training data, e.g., N pairs (x_i, y_i) of inputs x_i and observations y_i
- Training the model means finding parameters θ^* such that $f(x_i, \theta^*) \approx y_i$
- Define a loss function, e.g., $\sum_{i=1}^{N}(y_i - f(x_i, \theta))^2$, which we want to optimize
- Typically: Optimization based on some form of gradient descent
- **Differentiation** required



Types of Differentiation

1. Scalar differentiation: $f: \mathbb{R} \rightarrow \mathbb{R}$

$$y \in \mathbb{R} \; w.r.t. \; x \in \mathbb{R}$$

2. Multivariate case: $f: \mathbb{R}^N \rightarrow \mathbb{R}$

$$y \in \mathbb{R} \; w.r.t. \; vector \; x \in \mathbb{R}^N$$

3. Vector fields: $f: \mathbb{R}^N \rightarrow \mathbb{R}M$

$$vector \; y \in \mathbb{R}^M \; w.r.t. \; vector \; x \in \mathbb{R}^N$$

4. General derivatives:$f: \mathbb{R}^{MxN} \rightarrow \mathbb{R}PxQ$

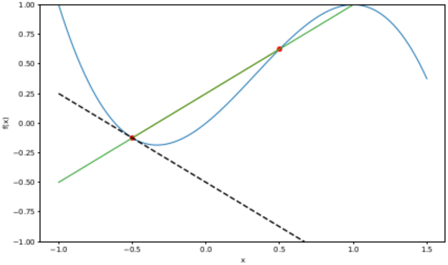
$$matrix \; y \in \mathbb{R}^{PxQ} \; w.r.t.matrix \; x \in \mathbb{R}^{MxN}$$

Scalar differentiation: $f : \mathbb{R} \to \mathbb{R}$

- Derivative defined as the limit of the difference quotient

$$f'(x) = \frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- Slope of the secant line through $f(x)$ and $f(x+h)$



Some Examples

$$
\begin{aligned}
f(x) &= x^n & f'(x) &= nx^{n-1} \\
f(x) &= sin(x) & f'(x) &= cos(x) \\
f(x) &= tanh(x) & f'(x) &= 1 - tanh^2(x) \\
f(x) &= exp(x) & f'(x) &= exp(x) \\
f(x) &= log(x) & f'(x) &= \frac{1}{x}
\end{aligned}
$$

Rules

- Sum Rule

$$(f(x) + g(x))' = f'(x) + g'(x) = \frac{df(x)}{dx} + \frac{dg(x)}{dx}$$

Rules

- Sum Rule

$$(f(x)+g(x))' = f'(x)+g'(x) = \frac{df(x)}{dx} + \frac{dg(x)}{dx}$$

- Product Rule

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x) = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx}$$

Rules

- Sum Rule

$$(f(x)+g(x))' = f'(x)+g'(x) = \frac{df(x)}{dx} + \frac{dg(x)}{dx}$$

- Product Rule

$$(f(x)g(x))' = f'(x)g(x)+f(x)g'(x) = \frac{df(x)}{dx}g(x)+f(x)\frac{dg(x)}{dx}$$

- Chain Rule

$$(g \circ f)'(x) = (g(f(x)))' = g'(f(x))f'(x) = \frac{dg(f(x))}{df} + \frac{df(x)}{dx}$$

- Quotient Rule

$$(\frac{f(x)}{g(x)})' = \frac{f(x)'g(x)-f(x)g(x)'}{(g(x))^2} = \frac{\frac{df}{dx}g(x)-f(x)\frac{dg}{dx}}{(g(x))^2}$$

Example: Scalar Chain Rule

$$(g \circ f)'(x) = (g(f(x)))' = g'(f(x))f'(x) = \frac{dg(f(x))}{df} + \frac{df(x)}{dx}$$

Beginner

$g(z) = 6z + 3$

$z = f(x) = -2x + 5$

$(g \circ f)'(x) = \underbrace{(6)}_{\frac{dg}{df}} \underbrace{(-2)}_{\frac{df}{dx}}$

$= -12$

Advanced

$g(z) = tanh(z)$

$z = f(x) = x^n$

$(g \circ f)'(x) = \underbrace{(1 - tanh^2(x^n))}_{\frac{dg}{df}} \underbrace{nx^{n-1}}_{\frac{df}{dx}}$