(Byte Sized) Machine Learning Lecture Notes

Simplified and with examples

Dr. Karthik Mohan

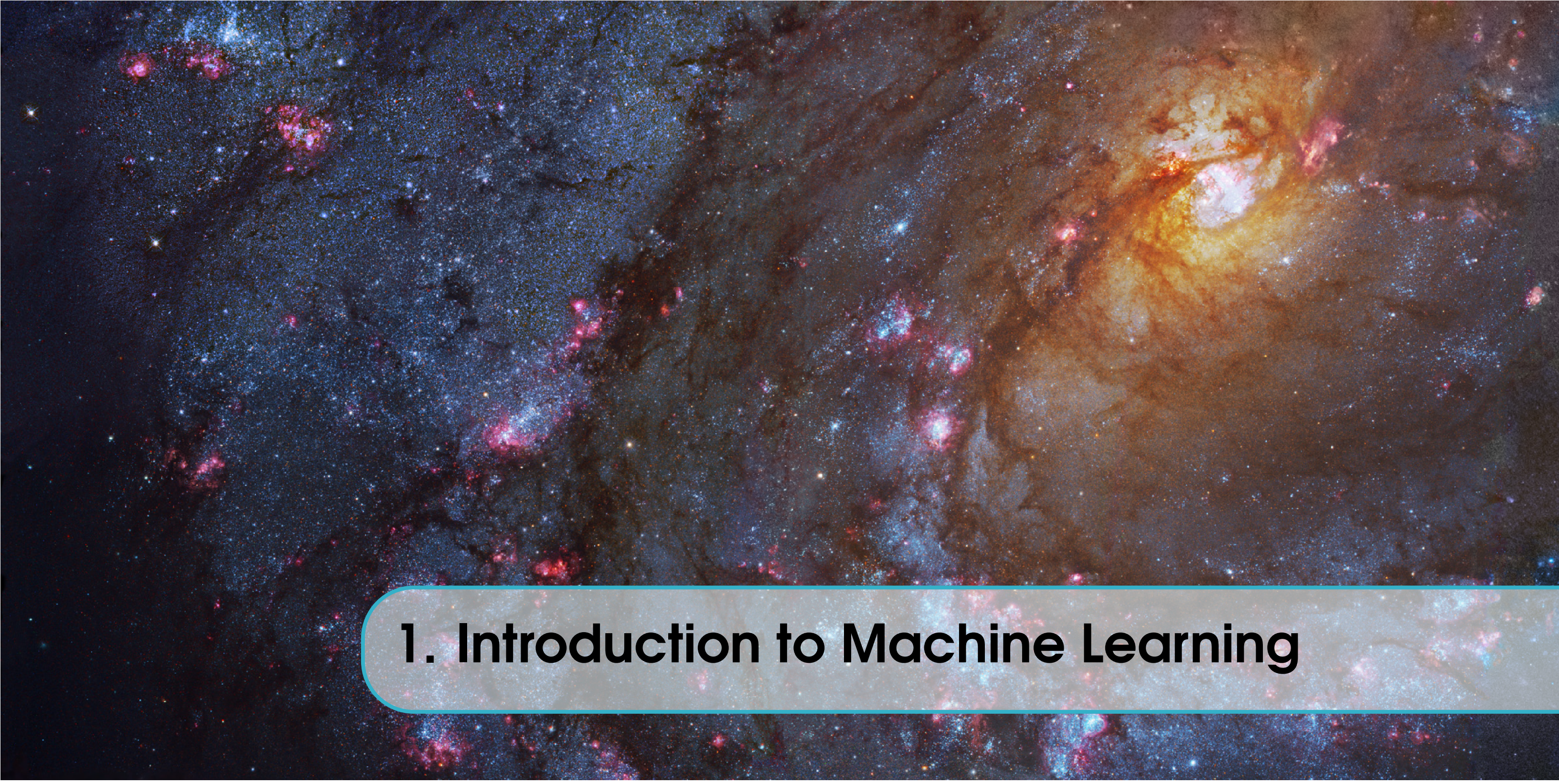
ADVANCED INTRO TO ML, LECTURE NOTES

HTTPS://BYTESIZEML.GITHUB.IO/ML2023/

These lecture notes are part of the course taught at University of Washington, Seattle in Winter 2023.

First release, January 2023

Contents



1. Introduction to Machine Learning

1.1 High-level Motivation

Data is in abundance in today's world. Every leading tech company is focused or moving towards data-driven decision making. Machine Learning in the real-world, bridges the gap between data and making data-driven decisions.

1.2 Motivating Example

Consider that you want to teach a kid to recognize objects, and specifically to recognize apples.

1. So you show the kid a red apple and say 'This is an apple.' The kid picks up on the color red and learns to associate it with apple.
2. You then show the kid a 'red cube' and ask, 'Is this an apple?' - The kid says, 'Yes that's an apple, because it is red in color.'
3. You then clarify, that an apple has to be circular in shape.
4. You then show the kid a red circular ball, and ask, 'Is this an apple?' The kid says, 'Yes, because it is red and circular in shape.'
5. You clarify again, that apples are circular but with a depression at the top.
6. The kid, by now, has understood, that an apple has to be circular in shape but with a depression at the top, and also red in color.
7. With 4 examples, the kid has quickly learned to recognize any apple in the wild!
8. To test the kid's understanding, you show the kid a green apple, and ask if it's an apple? The kid responds, 'No! It's not an apple because it's green in color!'
9. You shake your head and decide you will clarify this to the kid once you finish a business meeting that's about to start in a minute!

1.3 What is Machine Learning?

Here are three distinct but complementary defenitions of machine learning:

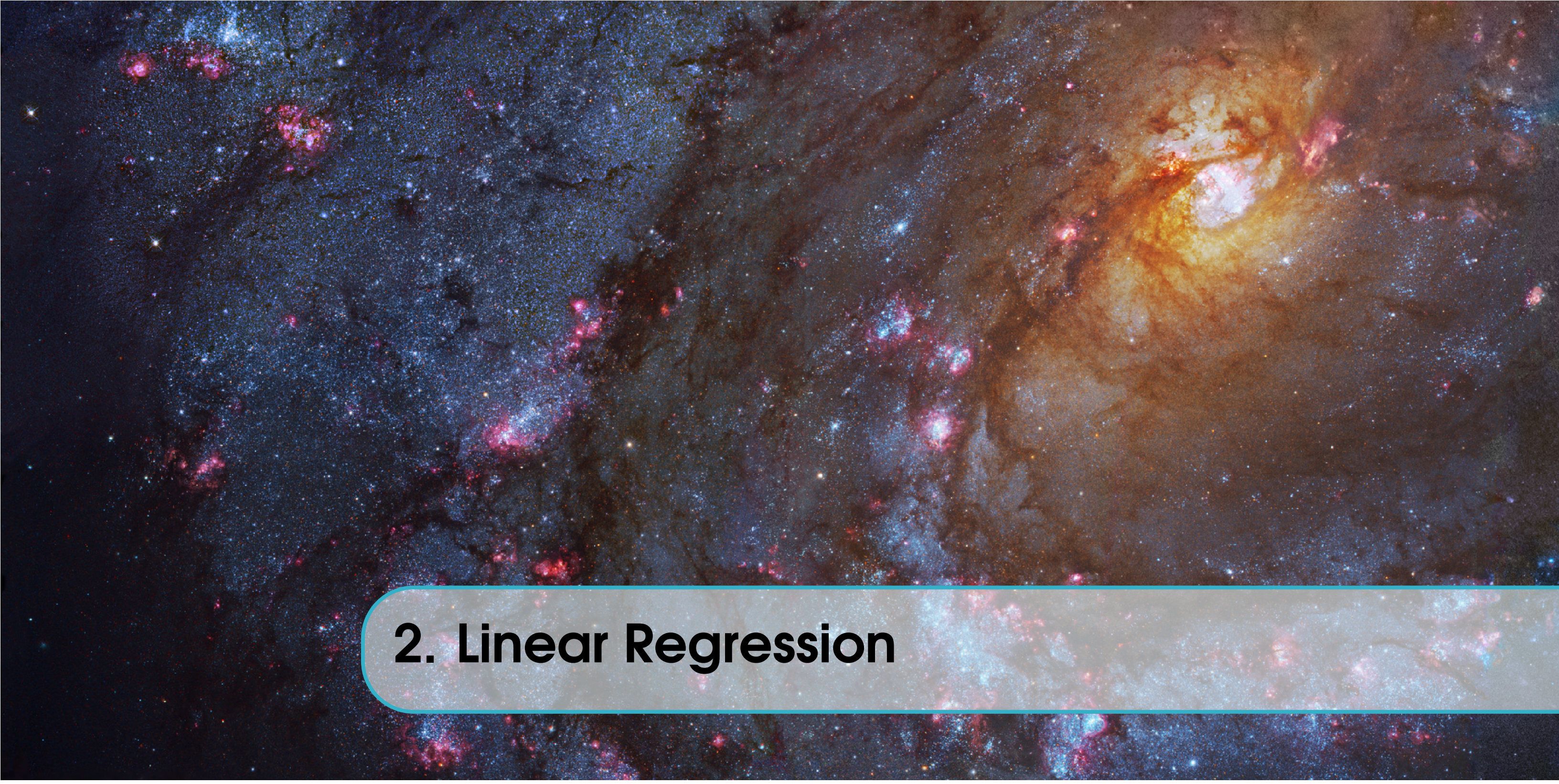
1. Machine Learning is a set of methods, tools, algorithms and frameworks to help us solve real-world problems with data.
2. Machine Learning are tools to learn useful patterns from data sets, which can then be used for decision making in the real-world
3. Machine Learning helps machines learn from data to interact with the real world as humans would (this definition is closer to that of Artifical Intelligence)

1.4 Methods used in Machine Learning

Methods in machine learning can be put in 3 broad categories: Supervised Learning, Unsupervised Learning and Reinforcement Learning. In this course, we focus extensively on the first two, which can be used to solve a majority of business problems that need machine learning.

1.5 Exercises

1. What phenomenon would help us understand why the kid couldn't identify a green apple as an apple and instead mistakenly identify it as a lime?



2. Linear Regression

2.1 Motivating Example

Consider that you want to predict the weight of a person, given their height! How would you do it? Is it generally true that taller a person is, the more they weigh? On an average, you might answer affirmative to this question. However, there are people who are tall and weigh less than average. Conversely, there are people who are short who weight more than average. But on an average, maybe we can derive a relationship between weight and height of a person!

2.2 What is Linear Regression?

In the above example, if the relationship we derive was 'linear' - We get linear regression! Regression is essentially a way to explain an output (weight in this example), as a function of the inputs (height in this example). What's an example of a linear function?

$$f(w) = wx + c$$

Here, f as a function of w is linear in w. w could represent a weighting term, x could be the height of the person and $f(w)$ predicts the height! What if now, instead of just using height to predict the weight, you also use BMI (body mass index) and perhaps resting heart rate (a measure of fitness)? We might get a more accurate fit. Now x is no longer just a scalar, it becomes a vector $\mathbf{x}$ and it represents height, BMI, and resting heart rate. Our output, prediction, is still the weight. So this changes, the linear regression model:

$$f(\mathbf{w}) = \mathbf{w}^T \mathbf{x} + c$$

2.3 Methodology behind Linear Regression

You can fit any line through the data and perhaps it could help explain the relationship between the height and weight of a person. And perhaps, it won't! But the *line of best fit*, would certainly *best* *explain* this relationship between the input and the output. Why? Because it is the line of best fit. Consider Figure 1 below. The red line represents the line of best fit. But look at all the other lines, they may explain the relationship between height and weight, but they certainly don't look like the best fit!

2.4 Delving into the Math behind Linear Regression

Let $\hat{y} = f(\mathbf{w})$ represent the **prediction** of the Linear Regression model. We want, the prediction to be as close to the ground truth y as possible for all possible data points. In other words,

$$\min_{\mathbf{w} \in \mathscr{R}^d} \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

also equivalent to,

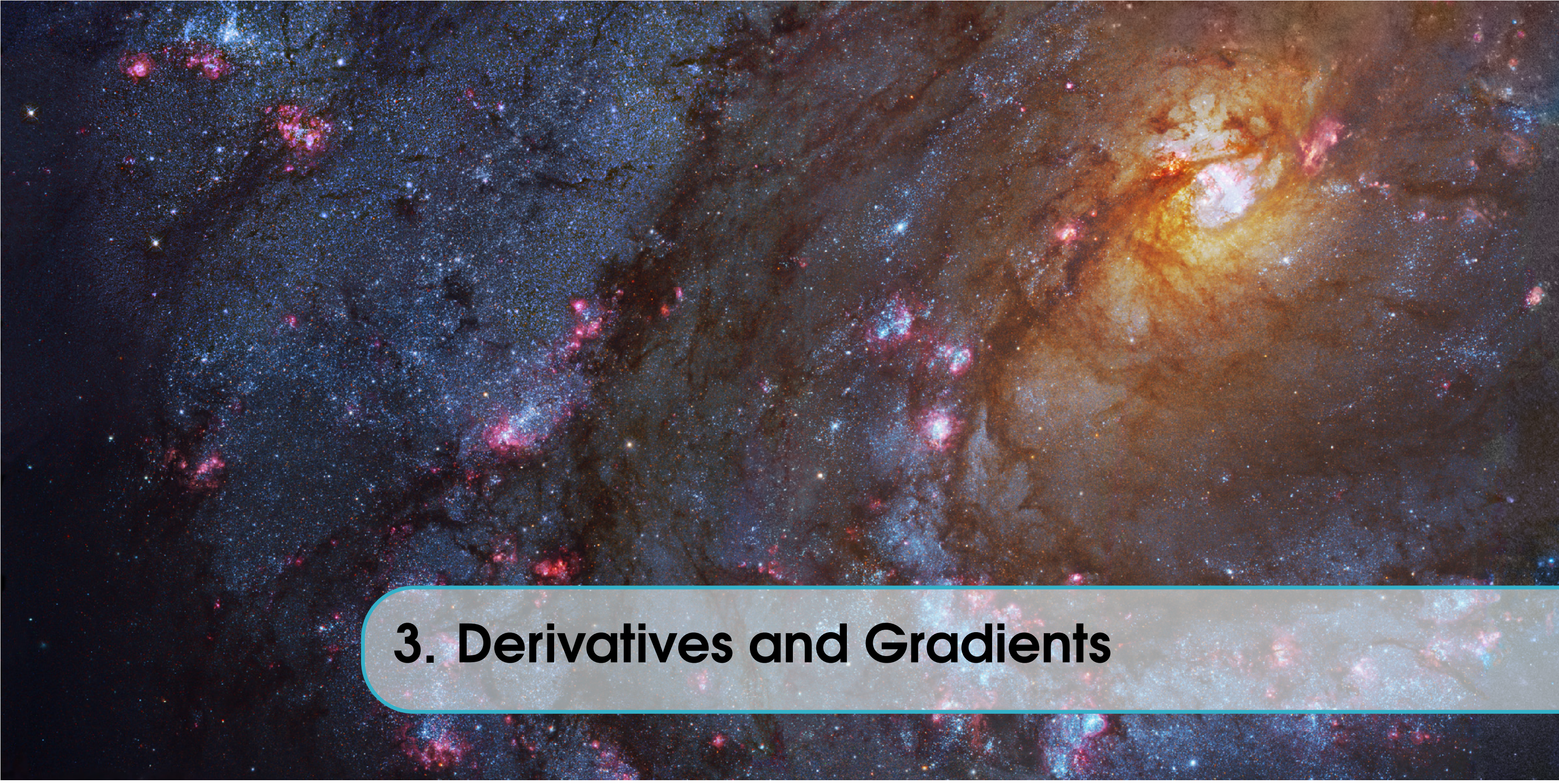
$$\min_{\mathbf{w} \in \mathscr{R}^d} \frac{1}{N} \sum_{i=1}^{N} (y_i - \mathbf{w}^T \mathbf{x}_i)^2$$

also equivalent to,

$$\min_{\mathbf{w} \in \mathscr{R}^d} \frac{1}{N} \|\mathbf{y} - X\mathbf{w}\|_2^2$$

where $\|\mathbf{z}\|_2^2 = \sum_i \mathbf{z}_i^2$ is the *Euclidean Norm* of $\mathbf{z}$.

2.5 Exercises



3. Derivatives and Gradients

3.1 Motivation

Believe it or not, the word, **gradient** is fundamental to most popular machine learning algorithms you will find on the market. Take any library, like *sci-kit learn* or a deep learning framework like *PyTorch*, under the hood - They use *gradients*, as a fundamental to learn from the data! Your favorite object detection method in ML, learns from data through *gradient descent*. So let's learn more on computing gradients.

3.2 What is a gradient?

Gradient is nothing but a collection of partial derivatives of a function. Derivative of a single variable function is something you would have encountered in a basic calculus course. For instance, what is the derivative of $f(w) = w^2$? It's $2w$. Now extrapolate derivative of function of single variable $w \in \mathbb{R}$ to a derivative of a function with respect to a vector, $\mathbf{w} \in \mathbb{R}^d$ and we get the *gradient*.

So what's the gradient of $f(\mathbf{w}) = \|\mathbf{w}\|_2^2$? It's $2\mathbf{w}$. The gradient of a function with respect to the vector has the same dimensions as the vector!

3.3 Hiking and Gradients!

Gradients have a cool property: At any given point in space, the gradient of a fucntion tells you the direction to move in, so you increase the function value in the fastest possible way! Imagine, you were to go hiking up a mountain. Usually, hiking trails have switchbacks - These are not the steepest ways to ascend the mountain. However, they are safer! Imagine, you shortcut the switchbacks and went straight up the mountain - Now, that would be the direction in which the gradient would be pointing to!