

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Mathematics . . . . .	4
1.1.1	Vectors and Matrices: Conventions . . . . .	4
1.1.2	Eigenvalues . . . . .	4
1.1.3	Gradient Descent . . . . .	4
1.1.4	Manifolds . . . . .	4
1.2	Statistics . . . . .	4
1.2.1	Probability and Random Variables . . . . .	4
1.2.2	Frequentist Statistics . . . . .	4
1.2.3	Bayesian Statistics . . . . .	4
1.3	Computation . . . . .	4
1.3.1	Numerics . . . . .	4
1.3.2	High Performance Computing . . . . .	4
1.3.3	Data and Data Representation . . . . .	4
1.3.4	Julia and Fluxes . . . . .	4
<b>2</b>	<b>Machine Learning</b>	<b>5</b>
2.1	Supervised Learning . . . . .	5
2.1.1	Loss Functions . . . . .	5
2.1.2	Classification and Regression . . . . .	5
<b>3</b>	<b>Feed-Forward: Multilayer Peceptrons and Deep Learning</b>	<b>7</b>
3.0.1	MLP . . . . .	7
3.0.2	Convolutional Network . . . . .	7
3.0.3	Deep Convolutional Network . . . . .	7
<b>4</b>	<b>Feed-Back: Recurrent Neural Networks</b>	<b>9</b>



# Chapter 1

## Introduction

The course is introduced by training a deep-convolutional neural network on an image classification task: birds or big cats. The training data will be shown as images of lions, leopards, and various types of corvids (black feathered birds). The validation set is some trivial examples and a panther. For each output the net will assign a probability and the panther as a dark example of a big cat will be on the boundary of the discriminator - a good example.

## 1.1 Mathematics

### 1.1.1 Vectors and Matrices: Conventions

### 1.1.2 Eigenvalues

### 1.1.3 Gradient Descent

### 1.1.4 Manifolds

## 1.2 Statistics

A statistical model is a function with a representation of data and a parametrisation. Usually, the data is used to inform this parametrisation. The function is termed a *distribution* and captures, in essence, the probability of a given data point by drawn at random from the data pool. What is the probability that this image is a bird?

### 1.2.1 Probability and Random Variables

### 1.2.2 Frequentist Statistics

### 1.2.3 Bayesian Statistics

Priors

Posteriors

## 1.3 Computation

Statistics (and in fact much of modern mathematics) requires the processing of large amount of data typically represented as some form of number. Humans are generally pretty good at understanding the concepts behind this processing, but not so good at doing the processing themselves. I am truly terrible. Computers and computer science have been developed to facilitate this task. We generally don't want to think too deeply about the lower level implementations when designing our machine learning tasks because we will get bogged down in the details. To this end many packages and ecosystems have been developed to abstract away these details but we should still be aware of them lest we encounter a grizzly pitfall.

### 1.3.1 Numerics

Integer Underflow and the Log Trick

### 1.3.2 High Performance Computing

### 1.3.3 Data and Data Representation

### 1.3.4 Julia and Flux

# Chapter 2

## Machine Learning

A simple definition: machine learning at its core is distribution matching: given some data and a canonical model representation learn parameters such that the model distribution matches the data. The models are generally classified as generative and discriminative. The methods for training a model can be split into: supervised, unsupervised, and reinforcement. We will focus on supervised discriminative models.

### 2.1 Supervised Learning

#### 2.1.1 Loss Functions

Usually specified as some function of the difference between the empirical distribution (observed from data) and the model distribution. A common choice is the mean-squared error.

#### 2.1.2 Classification and Regression

**Training**

**Linear Regression**

**Logistic Regression**

Why am I talking about this? Introduce the notion of a probability assigned to a class: is this a bird, or a cat. Logistic regression allows you to squash the continuum into the 0-1 range allowing you to assign it a probability. The classification task can then be thought of as: given a data point and a number of classes perform logistic regression on the data

**The Perceptron**

**The XOR Gate**

End the second lecture here.



# Chapter 3

## Feed-Forward: Multilayer Peceptrons and Deep Learning

3.0.1 MLP

3.0.2 Convolutional Network

3.0.3 Deep Convolutional Network





## Chapter 4

# Feed-Back: Recurrent Neural Networks