MPhil Data Intensive Science M1 Contents

Zhimei Liu

Michaelmas 2024

This document provides an overview of the MPhil course *Machine Learning for Data Intensive Science* in Michaelmas 2024, lectured by *Dr. Sven Krippendorf*.

Contents

1 Machine Learning Concepts **1**
1.1 Bias-Variance Tradeoff . . . 1

2 Optimisation Algorithms **1**
2.1 Gradient Descent . . . 1
2.1.1 Gradient Descent vs Newton's Method . . . 1
2.2 Stochastic Gradient Descent . . . 1
2.2.1 SGD with mini-batches . . . 1
2.2.2 Adding Momentum . . . 1
2.3 RMS-Prop . . . 1
2.4 ADAM . . . 1

3 Data **1**

4 Loss **1**
4.1 KL Divergence . . . 1
4.2 Linear regression loss . . . 1
4.2.1 Ridge regression . . . 1
4.2.2 Lasso regression . . . 1
4.2.3 Maximum A Posteriori Estimation (MAP) . . . 1
4.2.4 Example: Ising Model . . . 1
4.3 Logistic regression and Cross-entropy loss . . . 1
4.3.1 Binary Cross Entropy . . . 1
4.3.2 Categorical Cross Entropy . . . 1

5 Neural Network **1**
5.1 Perceptron . . . 1
5.2 Feed-Forward Neural Network . . . 1
5.3 Backpropagation . . . 1

6 Supervised Learning Models **1**

6.1 Support Vector Machine 1
6.2 Decision Tree 1
6.3 Ensemble Learning 1
6.3.1 Bagging 1
6.3.2 Boosting 1
6.3.3 Random Forests 1

7 Discrete Latent Variables **1**
7.1 K-means clustering (EM) 1
7.2 Gaussian Mixture Model (GMM) 1

8 Continuous Latent Variables/Dimensionality Reduction **1**
8.1 PCA 1
8.2 Autoencoder 1
8.3 T-SNE 1

9 Energy-based models **1**
9.1 Logistic regression as an energy-based model 1
9.2 Restricted Boltzman Machine (RBM) 1
9.3 Variational free energy 1
9.4 Hopfield Networks 1