CMSC472: Introduction to Deep Learning

James Zhang*

January 30, 2024

These are my notes for UMD's CMSC472: Introduction to Deep Learning, which is an elective ("live-TEX"-ed). This course is taught by Assistant Professor Abhinav Shrivastava.

Contents

*Email: jzhang72@terpmail.umd.edu

§1 Introduction to Statistical Learning I

The main idea of statistical learning is to just make sense of data.

Definition 1.1. **Supervised learning**: given inputs (data-label pairs), learn a model to predict output. In mathematical terms, we want to learn a prediction function f.

Definition 1.2. **Training (or learning)**: given a training set of labeled examples $\{(x_1, y_1), \cdots, (x_N, y_N)\}$, instantiate a predictor$f$. $y = f(x) \in \mathcal{H}$, this is known as the **hypothesis space**.

Definition 1.3. **Testing (or inference)**: apply f to a new test example and return the output $f(x)$. Note that we measure correctness using loss functions.

Note 1.4. Training and testing data should be i.i.d. (independent and identically distributed) samples from the same distribution D.

§1.1 Simple Models: Classification and Regression

Definition 1.5. A **nearest neighbor classifer** returns $f(x)$ = the label of the training example nearest to x. All we need is a distance function. Note that this model requires no training.

Similarly, a **K-Nearest Neighbor** finds k nearest points and $f(x)$ = vote for class labels with labels of the k nearest points. Note that kNN is more robust to outliers.

- kNN pros: Simple to implement, decision boundaries not necessary linear, works for any number of classes, nonparametric method
- kNN cons: need good distance function, slow at test time

Definition 1.6. A **linear classifier** is a linear function that separates the classes such that $f(x) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$.

Note 1.7. Perceptrons cannot do nonlinearly separated data.

- Linear pros: low-dimensional parametric representation, easy to learn, very fast at test time
- Linear cons: works for two classes(?), how to train the linear function, data is not linearly separable