Machine Learning - Notes

Dom Hutchinson

October 2, 2019

Contents

General

Lecturer - Carl Henrik Ek
Course Website - `http://carlhenrik.com/COMS30007/`
Course Repo - `https://github.com/carlhenrikek/COMS30007`
Course Subreddit - `https://www.reddit.com/r/coms30007/`

1 Introduction

Definition 1.1 - *Deductive Reasoning*
A method of reasoning in which the premieses are viewed as supplying all the evidence for the truth of the conclusion.

Definition 1.2 - *Inductive Reasoning*
A method of reasoning in which the premieses are viewed as supplying some evidence for the truth of the conclusion, rather than all the evidence. This allows for the conclusion of the *Inductive Reasoning* to be false.

Remark 1.1 - *Free-Lunch Theorem*
There are infinite number of hypotheses that perfectly explain the data. Adding a data point removes an infinite number of possibilities, but still leaves infinite possibilities.

Remark 1.2 - *The Task of Machine Learning*
When proposing to use machine learning on a task, one should consider the following questions:

i) How can we formulate beliefes ad assumptions mathematically?

ii) How can we connect our assumptions with data?

iii) How can we update our beliefs?

1.1 Probability Theory

Definition 1.3 - *Stochastic/Random Variable*
A variable whose value depends on outcomes of random phenomona.
e.g. $x \sim \mathcal{N}(0, 1)$.

Definition 1.4 - *Probability Measure,* $\mathbb{P}$
A function with signature $\mathbb{P} : \mathcal{F} \to [0, 1]$, where $\mathcal{F}$ is a sample space of rv X, and fulfils $\int_{-\infty}^{\infty} \mathbb{P}(x)dx = 1$.

Definition 1.5 - *Joint Probability Distribution*
A *Probability Measure* for multiple variables, $\mathbb{P} : X \times Y \to [0, 1]$.
Let n_{ij} be the number of outcomes where $X = x_i$ and $Y = y_j$ then

$$\mathbb{P}(X = x_i, Y = y_j) = \frac{n_{ij}}{\sum_{i,j} n_{ij}}$$

Definition 1.6 - *Marginal Probability Distribution*
A *Probability Measure* for one variable when the sample space is over multiple variables.
Let n_{ij} be the number of outcomes where $X = x_i$ and $Y = y_j$ then

$$\mathbb{P}(X = x_i) = \frac{\sum_j n_{ij}}{\sum_{i,j} n_{ij}}$$

Definition 1.7 - *Conditional Probability Distribution*
A *Probability Measure* for a variable, given another variable has a defined value. Let n_{ij} be the number of outcomes where $X = x_i$ and $Y = y_j$ then

$$\mathbb{P}(Y = y_j | X = x_i) = \frac{n_{ij}}{\sum_j n_{ij}}$$

Theorem 1.1 - *Product Rule*
For random variables X & Y

$$\mathbb{P}(X=x_i,Y=Y_j)=\mathbb{P}(Y=y_j|X=x_i)\mathbb{P}(X=x_i)$$

Theorem 1.2 - *Sum Rule*
For random variables X & Y

$$\mathbb{P}(X=x)=\sum_j \mathbb{P}(X=x,Y=y_j)$$

Theorem 1.3 - *Baye's Theorem*
For random variables X & Y

$$\mathbb{P}(X=x_i|Y=y_j)=\frac{\mathbb{P}(Y=y_j|X=x_i)\mathbb{P}(X=x_i)}{\mathbb{P}(Y=y_j)}$$

Definition 1.8 - *Expectaction Value,* $\mathbb{E}$
The mean value a random variable will produce from a large number of samples.

Continuous	Discrete
$\mathbb{E}(X)=\int_{-\infty}^{\infty}x\mathbb{P}(X)dx$	$\mathbb{E}(X)=\sum_{-\infty}^{\infty}x\mathbb{P}(X)dx$
$\mathbb{E}(f(X))=\int_{-\infty}^{\infty}f(x)\mathbb{P}(X)dx$	$\mathbb{E}(f(X))=\sum_{-\infty}^{\infty}f(x)\mathbb{P}(X)dx$

Definition 1.9 - *Variance*
Describes the amount of spread in the values a single random variable will produce.

$$\text{var}(X)=\mathbb{E}\left(x-\mathbb{E}(x))^2\right)=\mathbb{E}(X^2)-\bigg(\mathbb{E}(X)\bigg)^2$$

Definition 1.10 - *Covariance*
Describes the joint variability between two random variables.

$$\text{cov}(X,Y)=\mathbb{E}\bigg(\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\bigg)$$

Definition 1.11 - *Marginalisation*
The process of summing out the probability of one random variable using its joing probability with another rando variable.

$$\begin{array}{rrcl} \text{Continuous} & \mathbb{P}(X=x) & = & \int(X=x,Y=y)dy \\ \text{Discrete} & \mathbb{P}(X=x) & = & \sum_i \mathbb{P}(X=x,Y=y_i) \end{array}$$