Maths for ML 6.1

6 Probability & distributions

Use probability to measure the chance of something occurring in an experiment

Quantify uncertainty requires the idea of a Random variable

This a function which maps outcomes of an experiment to a set of properties that we are interested in

Associated w/ the RV is a function which measures the prob that an outcome will occur
- called probability dist

3 concepts that define prob space:
Sample space, the events, Prob of an event

6.1 Construction of a Prob Space

6.1.1 Philosophy

two interpretation of Probability:
Bayesian & frequentist

Bayesian uses probabilities to specify the degree of uncertainty the user has about an event
— Subjective Prob — Degree of belief

frequentist considers the relative frequencies of events to the total num of events — Prob defined as freq of event when data is unlimited

6.1.2. Prob & Random variables

3 ideas:

1. Probability Space - quantify the idea of a probability. Do not generally work directly w/ this space

2. Random variables - transfer the probs to a more convenient numerical space

3. Distribution or law assoc w/ RV

Modern prob is based on axioms of 3 concept space: Sample, event & probability measure

Prob space models real world process (known as Experiment) w/ random outcomes

The sample Space Ω

Set of all possible outcomes of the ~~exp~~ experiment

e.g. coin toss have a sample spaces

$\{hh, tt, ht, th\}$

the event Space $\mathcal{A}$

~~for each event A~~

= the space of Potential results of Experi

A subset of the sample space

= the collect of subsets of Ω

the Probability P

- with each event $A \in \mathcal{A}$
- we assoc ~~an~~ a num $P(A)$ that measures

Proba that event will occur

Putting these together

Prob of single event must lie in the int $[0,1]$

total prob of all outcomes in the Sample Space must be 1 (Ω)

Given a probability space (Ω)
we want to use to model a real-world phenom

In ML we often aviod explicitly referring to the probability space, instead refer to the probabilities on quantities of interest denoted as T

What does aviodiing prob space mean?

in trad prob theory - you start be def the Sample Space Ω e.g. Heads/tail over 2
~~{HH, TT, HT, TH}~~ H or T = $P(H) = 0.5$ $P(T) = 0.5 = 1$

ML focuses on the RV that model outputs or the direct probabilistic outputs

These are the quantities of interest

In the real world the sample space Ω is complex & vast - often imposs to define all possible outcomes

Instead focus on modelling to make accurate predictions

We care about these outcomes, not the underlying theoretic structure of the space

log reg models the cond prob $P(Y|X)$

Model params are learned from data using tech such as MLE

Rely on prob assumption but dont require knowledge of the prob space

This book T = target space and elements of T as states

introduce ~~ad~~ a function $X : \Omega \to T$

takes outcome & transforms into element of target space
- Quantity of interest

This mapping is the Random Var $X(\cdot)$

e.g Coin, 2 throws : HH, TT, HT, TH
count num of Heads 2 0 1 1
$T = \{0, 1, 2\}$ ← we want Probs of Thes

For a finite sample space Ω & finite T

- the function of an RV is a lookup fn
- for any subset $S \subseteq T$ we assoc a $P(S) \in [0, 1]$

Example of terminologies

Bag w/ coins: £, $

Draw 2 = 4 outcomes Ω = ££, $$, $£, £$

Composition of Bag = $ = 0.3

the event we want to know is dollar samples

RV = Ω to T = num of dollar draws

$X(\$\$) = 2$, $X(££) = 0$, $X(£\$) = 1$, $X(\$£) = 1$

$T = \{0, 1, 2\}$

2 draws are indep of each other

$P(x=2) = P(\$\$) = P(\$) \cdot P(\$) = 0.3 \cdot 0.3 = 0.09$

$P(x=1) = P(\$£) \cup P(£\$) = P(£\$) + P(\$£) =$

$\underbrace{0.3 \cdot (1 - 0.3)}_{\$£} + \underbrace{(1 - .3) \cdot 0.3}_{£\$} = 0.42$

$P(x=0) = P((£, £)) = P(£) \cdot P(£) = (1-0.3) \cdot (1-0.3) = 0.49$

Sample space also called event space