8 Machine Learning and Real-world Data (sht25)

You are a 22nd century historian researching the "FEE" (First Epidemic Era) of 2019–2025, for which records are patchy. You research which government policy was in place in any given week during this historic phase. Policies, in order of severity, are: No restrictions, Tier 1, Tier 2, Tier 3, and Lockdown.

(*a*) From other historic sources, you know the following about sequences of policy levels: if you are in a given policy level, there is a 40% chance you will stay there, a 20% chance that you will be upgraded to the next-highest (more severe) level next week, and a 10% chance that you will be downgraded to the next-lowest (less severe) policy level. The background lockdown probability (which applies if nothing more informative is known about lockdown) is 10%. For each observation sequence, there is also a 5% chance of the sequence ending at any point. Transitions to any other policy level beyond those already described are equally likely. Observation sequences begin with each policy level at equal likelihood.

Using the information given above, construct the full transition probability table.

[7 marks]

(*b*) You want to estimate which policy was in place for the first six weeks of 2025, but unfortunately, the only information you have about this is a sequence of Covid case numbers for these six weeks:

$[0-99],\ [0-99],\ [>200],\ [>200],\ [>200],\ [100-199].$

You know that case loads are associated with policy levels as in the Table below. Describe how you can calculate the sequence of most likely policy levels for these 6 weeks, giving numbers for at least three steps of the calculation. Assume that all policies are equally likely in the week preceding the first week. [8 marks]

	No Restriction	Tier 1	Tier 2	Tier 3	Lockdown
0-99 cases	5%	10%	20%	50%	90%
100-199 cases	15%	40%	40%	30%	9%
> 200 cases	80%	50%	20%	20%	1%

(*c*) In which respects is the modelling described above not fully adequate to describe an actual epidemic? [5 marks]