TH Quiz 1 - Solution

Prequrises + Nonparametric Bayesian

Linear Algebra:

1. A positive definite matrix A can be defined as one for which the quadratic form

$$X^T AX$$

is positive for any real value of the vector X. Show that a necessary and sufficient condition for A to be positive definite is that all of the eigenvalues λ_i of A are positive.

Solution: Recall the definition of an eigenvalue λ (and an eigenvector v):

$$Av = \lambda v > 0$$

For a matrix to be positive definite, $X^\star AX > 0$ for all x . But if x is an eigenvector of A, then

$$v^\star v\lambda$$

Since $\|v\| = v^\star v$ is necessarily a positive number, in order for $v^\star Av$ to be greater than 0, λ must be greater than 0.

Stochastic Processes:

1. On a chessboard, determine the expected number of moves that it takes a knight, starting in one of four corners of the chessboard, to return to its initial position, assuming that each knight move is equally likely to be any of its legal moves at the time.

Solution: We can view each square on the chessboard as a vertex on a graph consisting of 64 vertices, and two vertices are connected by an edge if and only if a knight can move from one square to another by a single legal move.

Since knight can move to any other squares starting from a random square, then the graph is connected (i.e. every pair of vertices is connected by a path).

Now given a vertex i of the graph, let d_i denote the degree of the vertex, which is number of edges connected to the vertex. This is equivalent to number of possible moves that a knight can make at that vertex (square on chessboard). Since the knight moves randomly, transition probabilities from i to its neighbors is $\frac{1}{d_i}$.

Now since the chain is irreducible (since the graph is connected) the stationary distribution of the chain is unique. Let's call this distribution π. Now we claim the following:

Claim: Let π_j denote jth component of π. Then π_j is proportional to d_j.

Proof: Let I be the fuction on vertices of the graph such that $I(i) = 1$ if i is a neighbor of j, and $I(i) = 0$ otherwise. Then

$$d_j = \sum_i I(i) = \sum_i d_i \frac{I(i)}{d_i} = \sum_i d_i p_{ij}$$

where p_{ij} is the transition probability from i to j. Hence we have $dP = d$ where P is the transition matrix of the chain, and $d = (d1, ..., dj, ..., d64)$. Thus $\pi P = \pi \rightarrow$Claim

Therefore, it follows that after normalising we have

$$\pi_j = \frac{d_j}{\sum_i d_i}$$

Finally we recall the following theorem

Theorem If the chain is irreducible and positive recurrent, then

$$m_i = \frac{1}{\pi_i}$$

Where m_i is the mean return time of state i, and π is the unique stationary distribution. Thus if we call the corner vertex 1, we have

$$m_1 = \frac{1}{\pi_1}$$

You can check that $\sum_i d_i = 336$, and we have $d_1 = 2$ (at corner knight can make at most 2 legal moves. Therefore $\pi_1 = \frac{1}{168}$ and $m_1 = 168$

2. A transition matrix P is said to be *doubly stochastic* if each column sum is 1; that is, if

$$\sum_i P_{i,j} = 1, \text{ for all } j$$

If such a chain is irreducible and has states $1, \ldots, m$ find its stationary probabilities.

Solution: we calim that all othe indexes of the stationary matrix is equal to $\frac{1}{m+1}$ It is clear that the sum of each row and column is equal to 1, and each vertex can reach any other vertex, so it's irreducible and has only one returning class, so it is unique according to a theorem and the staitionary probabilities:

$$\forall j \sum^{i} \pi_i p_{i,j} = \pi_i$$

$$\sum_i \pi_i p_{i,j} = \sum_i \frac{1}{m+1} p_{i,j} = \frac{1}{m+1} \sum_i p_{i,j} = \frac{1}{m+1} = p_{i,j}$$

Non-Parametric Bayesian:

1. Given prior density $g(\mu)$ and observation $X \sim Poi(\mu)$, you compute $g(\mu|x)$, the posterior density of μ given x. Later you are told that x could only be observed if it were greater than 0. Does this change the posterior density of μ given x?
2. What is the difference between parameter and nonparametric models?

Solution: In a parametric model, the number of parameters is fixed with respect to the sample size. In a nonparametric model, the (effective) number of parameters can grow with the sample size.

Parametric models assume some finite set of parameters θ. Given the parameters, future predictions, x, are independent of the observed data, D:

$$P(x|\theta, D) = P(x|\theta)$$

therefore θ capture everything there is to know about the data.

So the complexity of the model is bounded even if the amount of data is unbounded. This makes them not very flexible.

Non-parametric models assume that the data distribution cannot be defined in terms of such a finite set of parameters. But they can often be defined by assuming an infinite dimensional θ. Usually we think of θ as a function.

The amount of information that θ can capture about the data D can grow as the amount of data grows. This makes them more flexible.