

Recuperatorio (30pts)

Programacion Estadística II

Lic. Alvaro Chirino Gutierrez

17/7/2020

Instrucciones

- Recuperatorio: Elija el parcial que va a recuperar y responda **únicamente** las preguntas de ese recuperatorio
 - Duración: 2 horas. 18:00 a 20:00
 - Forma de entrega, incluir solo las respuestas en archivos separados, el nombre del archivo debe ser: respuesta_1, respuesta_2.
 - La entrega al correo achirino.stat@gmail.com máximo hasta las 20:00, pasado el tiempo no se tomara en cuenta.
 - Incluir en el correo su nombre completo y parcial que esta recuperando
-

Primer Parcial

Pregunta 1 (15 pts.)

- PCA: Empleando la base de datos de las elecciones nacionales del 20 de octubre de 2019, seleccione las mesas para Bolivia y las elecciones para presidente.
 - Convierta los valores absolutos a valores relativos para los partidos políticos, incluya blancos y nulos.
 - Aplique el método de componentes principales y aplique el criterio de eigenvectores superiores a 1 para retener los componentes.
 - Calcule los componentes retenidos e inclúyalos en la base de datos
 - Genere un gráfico de histogramas para ver la distribución de los componentes retenidos

Pregunta 2 (15 pts.)

Con los componentes retenidos de la pregunta anterior, defina una clasificación de las mesas empleando el método k-center y el jerárquico, encuentre el valor óptimo de k entre 2 y 5 para las distancias euclideana y manhattan.

- Para el método k-center tomé como centro solo los medoides.
- Para el jerárquico tome en cuenta los tres enlaces.

Comente cuál es el valor óptimo de k y bajo que características.

Nota para ambas preguntas: Si tiene dificultades con el tamaño de las bases de datos en su computadora, realice un filtro y seleccione solo un departamento.

Segundo Parcial

Pregunta 1 (15 pts.)

Usando la encuesta a hogares 2018, realice de forma secuencial lo siguiente:

- Seleccione al jefe/jefa del hogar de hogares que tengan al menos una persona de 60 años o más
- Para la variable ingreso per cápita del hogar (en logaritmo natural) construya un modelo lineal con las siguientes variables dependientes:
 - edad del jefe
 - sexo del jefe del hogar
 - número de miembros
 - número de personas de 60 años o más
 - número de personas de 5 años o menos
 - número de personas entre 6 y 30 años
 - número de personas entre 31 y 59 años
 - máximo años de educación en el hogar
 - hogar con acceso a electricidad
 - servicio sanitario privado
 - hogar pobre/no pobre
- Realice los siguientes pasos sobre el modelo
 - Elimine a los hogares con un ingreso per capita mayor al percentil 98
 - Según la naturaleza de las variables defina el modelo lineal y elimine a las no significativas ($p < 0.05$)
 - Realice el test de normalidad sobre los errores y comente
 - Realice un test de multicolinealidad y comente
 - Realice un test de Homocedasticidad y comente
 - Incluya la interacción edad con número de miembros
 - Incluya los polinomios hasta el grado 3 para la variable edad

Pregunta 2 (15 pts.)

Usando la base de datos del ejercicio anterior, para la variable pobreza; elabore un modelo de clasificación usando los métodos:

- Logístico
- Árbol de clasificación (incluya el gráfico)
- Naive Bayes

Considere todas las variables para la clasificación en los distintos métodos.

¿Qué método es mejor?, justifique su respuesta.