IIMAS - UNAM

Paradigmas y Programación para Ciencia de Datos - Primer parcial

3

Nombres: Villala Pineda Luis Enrique Turing

Contesta:

- 1. Respecto al artículo Computing machinery and intelligence de Alan Turing, selecciona alguna de las visiones opuestas a la pregunta principal que no sea The Teological Objection y: (5p)
 - a) descríbela brevemente
 - b) argumenta si estás a favor o en contra de la misma
- 2. Respecto a la Agenda Nacional Mexicana de Inteligencia Artificial, elige alguna sección y contrasta las conclusiones mencionadas con el estado actual de desarrollo en esa área (5p)
- 3. Selecciona las dos características más importantes para el conjunto de datos Iris y con ellas obtén las funciones discriminantes de: (10p)
 - a) Un clasificador determinista a priori
 - b) Un clasificador estadístico a priori

Nota: Debe quedar clara la forma en la que elegiste las características para trabajar, así como argumentar el motivo por el cual decidiste usar esa forma.

- 4. Para la implementación propia del Perceptrón, simula el funcionamiento del método One versus All con ayuda del conjunto de datos Iris y evalúa cada modelo, puede ser con una métrica o una gráfica (10p)
- 5. Para el conjunto de datos de vinos: (15p)
 - a) Aplica la regresión logística al conjunto completo reducido y evalúa su rendimiento
 - b) Con ayuda de PCA, determina el número de componentes que conservan el 75 % de la información y realiza la reducción a ese número de componentes
 - c) Aplica la regresión logística al conjunto reducido con PCA y evalúa su rendimiento
 - d) Reduce al mismo número de componentes con ayuda de LDA
 - e) Aplica la regresión logística al conjunto reducido con LDA y evalúa su rendimiento

6. Elige:

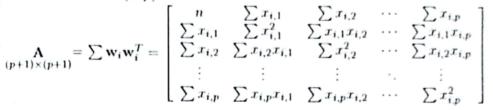
- a) Matriz de covarianzas
- b) Matriz de dispersión y escribe un código que obtenga la obtenga desde cero (10p)
- \mathcal{X} . Implementa desde cero KernelPCA con la siguiente función de kernel: $k\left(x^{(i)},x^{(j)}\right) = \alpha + \sum_{l=0}^{n-1} \left(x^{(i)} \cdot x^{(j)}\right)^{l}$

Después prueba su funcionamiento para separar círculos concéntricos; compara los resultados con el uso de PCA sobre los mismos datos (15p)

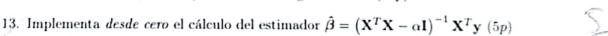
8. Elige dos estimadores y demuestra que definen algoritmos escalables (10p)

$$\overline{\mathbf{x}} = n^{-1} \begin{bmatrix} \sum_{i=1}^{n} x_{i,1}, \dots, \sum_{i=1}^{n} x_{i,p} \end{bmatrix}^{T}; \quad \hat{\sigma}_{j}^{2} = n^{-1} \sum_{i=1}^{n} x_{i,j}^{2} - \overline{x}_{j}^{2}; \quad \hat{\sigma}_{jk} = n^{-1} \sum_{i=1}^{n} x_{i,j} x_{i,k} - \overline{x}_{j} \overline{x}_{k}$$

9. Escribe un código que obtenga la matriz de momentos aumentada a partir de un conjunto de vectores de datos: (15p)



- 10. Obtén las expresiones para determinar los intervalos de contenedores para histogramas bidimensionales en función del número de contenedores (5p)
- 1). Dentro del paquete sklearn.datasets, existe un generador de datos para regresión llamado make_regression:
 - a) revisa en la documentación cómo generar un conjunto de datos con los siguientes parámetros: (15p)
 50 muestras de 7 características, ruido=20, un valor de tu elección en la semilla del generador de número aleatorios para replicar tus resultados
 - b) Separa los datos de entrenamiento y pruebas con una semilla establecida para replicar el experimento
 - c) Aplica regresión lineal al conjunto de entrenamiento
 - d) Evalúa el resultado sobre el mismo conjunto de entrenamiento y después sobre el de pruebas
 - 12. Con los datos del problema anterior, escribe un código para comparar el resultado de la regresión lineal estándar y el uso de las regularizaciones Ridge y ElasticNet con al menos dos valores diferentes para los parámetros de regularización (10p)



- 14. Descarga el archivo que se encuentra en https://bit.ly/3zRrMtH y posteriormente: (15p)
 - a) determina el grado del polinomio que mejor se ajusta a los datos, la primera columna se considera como dependiente y la segunda como independiente; evalúa su rendimiento
 - b) posteriormente aplica regresión lineal, obteniendo también su evaluación y comparándola con la obtenida en la regresión polinomial
- 15. Determina el mayor y menor valor (absoluto) decimal que puede representarse en punto flotante de simple precisión (5p)

Nota: Debes adjuntar el código correspondiente en todas las respuestas en que hayas escrito alguno



