

Parcial Primer 35 %

Edgar Andres Gutierrez Caceres

12 de marzo de 2018

1. Introducción

El reconocimiento de patrones es la ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos o abstractos, con el propósito de extraer información que permita establecer propiedades de entre conjuntos de dichos objetos.

El reconocimiento de patrones también llamado lectura de patrones, identificación de figuras y reconocimiento de formas consiste en el reconocimiento de patrones de señales. Los patrones se obtienen a partir de los procesos de segmentación, extracción de características y descripción donde cada objeto queda representado por una colección de descriptores. El sistema de reconocimiento debe asignar a cada objeto su categoría o clase (conjunto de entidades que comparten alguna característica que las diferencia del resto). Para poder reconocer los patrones se siguen los siguientes procesos:

- adquisición de datos
- extracción de características
- toma de decisiones

El punto esencial del reconocimiento de patrones es la clasificación: se quiere clasificar una señal dependiendo de sus características.

2. Objetivos del parcial

2.1. Objetivos

- Visualizar e interpretar datos
- Diseñar e implementar clasificadores bayesianos
- Utilizar el método de análisis de componentes principales

3. Conceptos previos

3.1. Clasificador Bayesiano

El enfoque probabilístico bayesiano provee un marco formal para construir clasificadores óptimos bajo ciertos criterios (como el minimizar el error de clasificación o el costo de una mala clasificación).

Sin embargo, si aplicamos el enfoque bayesiano en una forma directa, la complejidad computacional (en memoria y tiempo) crece exponencialmente con el número de atributos de los objetos. Una alternativa para enfrentar este problema, es el clasificador bayesiano simple, que asume que todos los atributos de un objeto son estadísticamente independientes dada la clase. Esto hace que la complejidad crezca linealmente con el número de atributos, pero puede hacer que la efectividad del clasificador decrezca si los atributos no son realmente independientes dada la clase.

3.2. PCA

En estadística, el análisis de componentes principales (en español ACP, en inglés, PCA) es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Técnicamente, el ACP busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. Esta convierte un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables sin correlación lineal llamadas componentes principales. El ACP se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos. El ACP comporta el cálculo de la descomposición en autovalores de la matriz de covarianza, normalmente tras centrar los datos en la media de cada atributo. Debe diferenciarse del análisis factorial con el que tiene similitudes formales y en el cual puede ser utilizado como un método de aproximación para la extracción de factores.

3.3. Preguntas

1. El conjunto de datos **data** contiene datos pertenecientes a 2 clases $\{a, b\}$
 - Grafique los datos utilizando un color distintivo para cada clase.
 - Determine el centro de cada clase $\{\mu_a, \mu_b\}$
 - Determine las matrices de covarianza de cada clase $\{\Sigma_a, \Sigma_b\}$. Que se puede concluir?
 - Determine los auto-vectores $\{v_1, v_2\}$ y los auto-valores $\{\lambda_1, \lambda_2\}$ de cada matriz de covarianza Σ
 - Determine y visualice el histograma de los datos.
2. El conjunto de datos **data2** contiene datos pertenecientes a 2 clases $\{a, b\}$
 - Divida los datos aleatoriamente en conjunto de **prueba** (20%) y de **entrenamiento** (80%).
 - Visualice el conjunto de entrenamiento con un color para cada clase.
 - Implemente un clasificador bayesiano gaussiano:
 - a) Estime la función de verosimilitud de cada clase.
 - b) Visualice la clasificación realizada sobre el conjunto de prueba.
 - c) Determine el error de clasificación sobre el conjunto de prueba.
 - Implemente un clasificador bayesiano gaussiano naive:
 - a) Estime la función de verosimilitud de cada clase.
 - b) Visualice la clasificación realizada sobre el conjunto de prueba.
 - c) Determine el error de clasificación sobre el conjunto de prueba.
3. Utilizando el conjunto de datos del Punto 2:
 - Determine la transformación de \mathbf{R}^3 a \mathbf{R}^2 utilizando PCA (sobre el conjunto de entrenamiento).
 - Implemente un clasificador bayesiano gaussiano sobre los datos transformados:
 - a) Estime la función de verosimilitud de cada clase.
 - b) Visualice la clasificación realizada sobre el conjunto de prueba.
 - c) Determine el error de clasificación sobre el conjunto de prueba.
4. Concluya sobre los resultados obtenidos de los 3 clasificadores construidos anteriormente.

3.4. Observaciones

Elabore un informe IEEE con el desarrollo del parcial y envíelo al campus virtual con todos los adjuntos en *.rar., De igual forma envíe una copia al correo edgar.gutierrez@usatunja.edu.co