Universidad de Salamanca
Grado en Matemáticas

REVISIÓN DE MÉTODOS MULTIVARIANTES SUPERVISADOS Y NO SUPERVISADOS

Trabajo Fin de Grado

Alumno: Pedro Ángel Fraile Manzano

Tutoras: Ana Belén Nieto Librero y Nerea González García

Salamanca, Julio de 2023

Índice general

Introducción

0.1. Distribuciones Multivariantes

Definición 0.1.1.

Capítulo 1

Análisis de Componentes Principales

1.1. Introducción

Dado un conjunto de p variables correladas que debemos examinar, hay problemas cuando p es especialmente alto. En aplicaciones como el campo del aprendizaje automático, si el número de variables estudiadas es mayor que el número de muestras recogidas, el modelo tenderá al overfitting y, por tanto, fallará al generalizar sus predicciones.
La base conceptual del análisis de componentes principales, PCA de ahora en adelante, es que en el caso de que tengamos un conjunto de variables correladas, entonces, dentro de las p habrá variables que dan la misma información y por tanto, podemos reducir la dimensionalidad de nuestro conjunto de datos.
El objetivo del análisis de componentes principales es comprobar y estudiar si combinaciones lineales de nuestras p variables iniciales pueden traer consigo la mayor variación de los datos.
Por último, el resultado obtenido tras aplicar el PCA, es un conjunto de variables no correladas. En principio, si el conjunto de las p variables son cerca de no estar correladas, no tiene sentido aplicar PCA, ya que este encontrará variables cercanas y ordenadas por varianza.

1.2. Definición y obtención de las Componentes Principales

Sea un vector aleatorio $\mathbf{X} = [X_1, \ldots, X_p]$ con media μ y matriz de covarianzas $\Sigma = Cov(X_i, X_j)$.

Llamaremos primera componente principal $\mathbf{Y}_1$ del vector aleatorio $\mathbf{X}$ a una combinación lineal de las p componentes:

$$\mathbf{Y}_1 = a_1^1\mathbf{X}_1 + \ldots a_p^1\mathbf{X}_p = \mathbf{a}_1^T\mathbf{X}$$

Donde el vector $\mathbf{a}_1 \in \mathbb{R}^p$ y además se cumple que $\mathbf{a}_1^T\mathbf{a}_1 = 1$, es decir, unitario para evitar multiplicar la norma del vector por una constante arbitraria y poder continuar comparando.

Este vector $\mathbf{Y}_1 = \mathbf{a}_1^T\mathbf{X}$ tiene como varianza $Var(\mathbf{Y}_1) = Var(\mathbf{a}_1^T\mathbf{X}) = \mathbf{a}_1^T\Sigma\mathbf{a}_1$. Esta es la función a maximizar, la cual está sometida a la restricción $\mathbf{a}_1^T\mathbf{a}_1 = 1$.

Para resolver este problema , tenemos que los puntos de una función diferencial como la hemos definido $Var(\mathbf{Y}_1)$