Matemáticas para Ciencia de Datos

Ejercicios 3

Sea A una matriz de datos dada por el archivo `dataset.csv`.

- Sospechamos que nuestra matriz de datos tiene columnas redundantes (se pueden poner como combinaciones lineales de las otras) por la naturaleza de su creación, ¿cuál es el máximo número de columnas linealmente independiente que puede tener A?

- Supón que quieres «borrar» de A las columnas que dependan linealmente de otras sin sacrificar la dimensión de su rango (su *rank*). Una primera búsqueda por el internet nos arroja la siguiente respuesta misteriosa muy relacionada aquí:

  > [...] One way to do what you want is to apply numpy.linalg.qr to the transpose, and check the non-zero components of the R matrix. The corresponding columns (in the transpose matrix, i.e., the rows in your original matrix) are independent.

  Es decir, pide considerar la descomposión de $A = QR$, aquí Q es una matriz ortogonal (columnas y renglones ortogonales entre sí) y R una matriz triangular superior con las entradas debajo de la diagonal principal son cero. ¿Es correcta esta respuesta? Justifica tu respuesta con un contra-ejemplo o da una serie de razonamientos para justificar que la respuesta es correcta.

- Elimina efectivamente de A las columnas que sean combinaciones lineales de otras. ¿Hay una sola manera de hacer esto? Identifica si te es posible las relaciones lineales de la columnas eliminadas respecto a las otras.

- Sea $\mathbb{1}$ el vector de solo unos en $\mathbb{R}^n$. Para cada $x \in \mathbb{R}^n$ sea $x_{\text{centered}} = x - \mu\mathbb{1}$ donde μ es el promedio de los elementos de x (*sample mean*). Responde las siguientes preguntas:
  1. Sea A_{centered} la matriz que resulta de centrar cada columna de A. ¿Qué efecto geométrico tiene sobre A este procedimiento?
  2. Responde si x y y son dos columnas de A linealmente independientes, ¿también lo son x_{centered} y y_{centered}?
  3. Responde si x y y son dos columnas de A ortogonales, ¿también lo son x_{centered} y y_{centered}?
  4. ¿Cuál es la media de cualquier vector x_{centered} ?
  5. Para cada x, encuentra el vector unitario con dirección x_{centered}

- Encuentra, si existe, la/una combinación lineal de las primeras 2 columnas «más cercana» a la columna número 6. Haz lo mismo pero con las primeras 3, 4 y 5 columnas. En cada caso decide si la respuesta existe y es única.