

Titulo
Giovanni Gamaliel López Padilla

Problema 2.1

Sea $\{x_i\}$ un conjunto de n vectores d dimensional. Definimos la matriz Kernel $[K_{i,j}]$ con $K_{i,j} = \langle x_i, x_j \rangle$ y \mathbb{D}^2 la matriz de distancias al cuadrada correspondiente. Verifica la identidad que usamos en clase:

$$\mathbb{D}^2 = c1^t + 1c^t - 2\mathbb{X}\mathbb{X}^t$$

con 1 un vector de unos de longitud n y c el vector de longitud n con elementos $(\mathbb{K}_{i,i})_{i=1}^n$

Problema 2.2

En la página 18 del archivo recpat6.pdf de la clase del 9 de febrero, verifica cómo que se obtiene la expresión $K_\Phi(x, y) = (1 + \langle x, y \rangle)^2$. De manera similar, supongamos que se define otro kernel K :

$$K(x, y) = \langle x, y \rangle^3, x, y \in \mathbb{R}^2$$

Busca una función $\Phi()$ tal que:

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

Problema 2.3

Sea S un conjunto finito. Definimos como medida de similitud entre dos subconjuntos A y B de S :

$$K(A, B) := \#(A \cap B)$$

Busca una función tal que:

$$K(A, B) = \langle \Phi(A), \Phi(B) \rangle$$

Problema 2.4

Decidimos que dos observaciones x_i, x_j son conectados por una arista en el grafo correspondiente si x_i está entre los k -vecinos más cercanos de x_i o x_j está entre los k -vecinos más cercanos de x_i . Muestra que la adición de una sola observación en

este ejemplo puede destruir por completo el desenrollamiento. Márcala en el dibujo y explícalo.

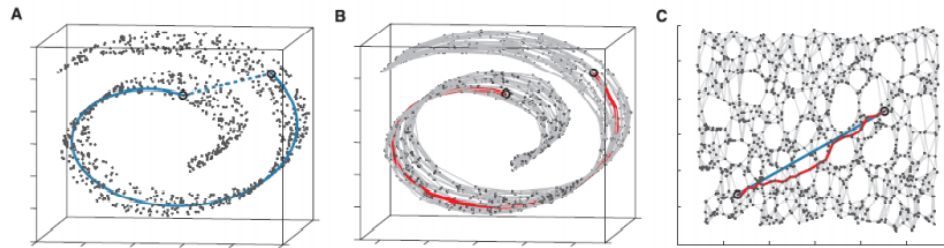


Figura 1

Problema 3.1

Trabajamos con los de datos fashion MNIST. Ver <https://www.kaggle.com/zalando-research/fashionmnist> Se trata de imágenes 28x28 de diez diferentes tipos de prendas. Trabajaremos con [fashion-mnist_train.csv](#)

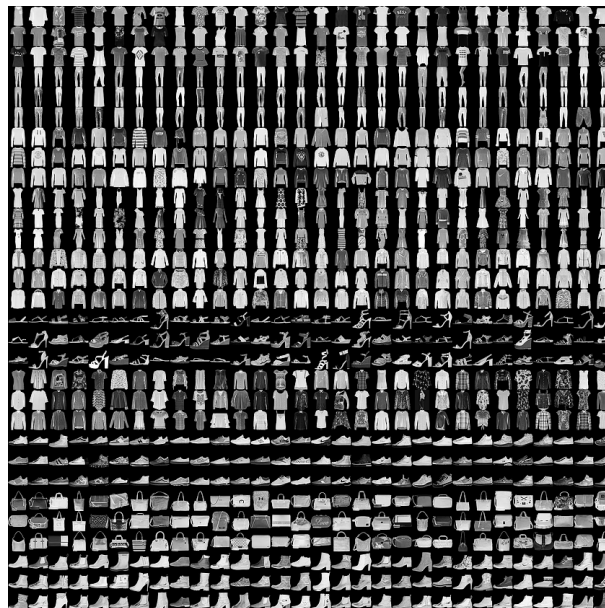


Figura 2

Busca visualizaciones 2D y 3D basadas en PCA de las imágenes de T-shirts (clase "0"). ¿ Ves posible encontrar interpretaciones de los componentes como lo hicimos en clase con la base mnist (clásico) de dígitos?

Problema 3.2

Trabajamos con datos de calificaciones de películas de Netflix por usuarios: <https://grouplens.org/datasets/movielens/latest/> Nos limitamos a la base chiquita. Busca algunas visualizaciones informativas de estos datos y coméntalos. Aplica MDS para obtener una visualización de las películas, explora diferentes kernels (basándose en

el vector de calificaciones de cada película y/o los géneros a los cuales cada película pertenece). Hay muchísimas calificaciones faltantes. Límitate a un subconjunto chiquito que se puede trabajar fácilmente.