1 Kernel Methods

Question 1: Say we have a linear dimensionality reduction technique. That is an algorithm that does whatever it does and in the end finds a projection matrix W of size $d \times K$ and computes $Y = XW$. Which of the following statements are true:

A. For the right algorithm, each column of W can be written as linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$

B. Algorithm can always be equivalently rewritten with a W such that each column of W is a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$

C. Y can sometimes be computed based on inner products between data points

D. Y can always be computed based on inner products between data points

2 t-SNE

Question 2: Given points $\mathbf{x}_1, \ldots, \mathbf{x}_n$, for any $t, s \in [n]$, let

$$p_{t \to s} = \frac{\exp(-\frac{\|\mathbf{x}_s - \mathbf{x}_t\|^2}{2\sigma^2})}{\sum_{u \neq t} \exp(-\frac{\|\mathbf{x}_u - \mathbf{x}_t\|^2}{2\sigma^2})}$$

Now define $P_{s,t} = \frac{p_{t \to s} + p_{s \to t}}{2n}$ and assume $P_{t,t} = 0$ for any t. Show that P is a valid probability distribution over $[n] \times [n]$.