

867002105_HW8

Charles Dotson

November 6, 2022

1 Q1

S is a symmetric and square covariance matrix. Thus, the eigenvalue of $S = \psi\theta\psi^{-1}$. Where, ψ is a matrix with the eigenvectors as the columns and θ has all eigenvalues along the diagonal with all other entries of the square matrix being zero. Thus,

$$\begin{aligned} J(u_1) &= u_1^T S u_1 \\ &= u_1^T \psi \theta \psi^{-1} u_1 \\ &= \theta \psi^{-1} u_1 \quad (\text{because } u_1 \text{ and } \psi \text{ are orthonormal}) \\ &= \theta \psi^T u_1 \quad (\text{because } \psi_1 \text{ are orthogonal}) \\ &= \left\langle \begin{bmatrix} \cdots \lambda_1 u_1 \cdots \\ \vdots \\ \cdots \lambda_k u_k \cdots \end{bmatrix}, u_1 \right\rangle \\ &= \lambda_1 u_1 u_1 + \dots + \lambda_k u_1 u_k \\ &= \lambda_1 \quad (\text{using orthonormality}) \end{aligned}$$

2 Q3

We start by proposing that a rank 1 matrix can always be written in the form of $A = xy^T$. By definition this is true, $S = xx^T$.

We can see if we multiply it by a vector w , $Sw = xx^Tw = (x^Tw)x$. With the term x^Tw always being a scalar value we see that if w is a eigenvector of S , it must be a multiple of x .

Finally we will show that for a matrix $S = xx^T$ to have a rank of 1, the $|S| = 0$. Thus,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_k \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_k \\ x_1x_2 & x_2^2 & \cdots & x_2x_k \\ \vdots & \vdots & \ddots & \vdots \\ x_1x_k & x_2x_k & \cdots & x_k^2 \end{bmatrix}$$

One can easily see that determinate of the above matrix is symmetric and square. Thus, call the determinet coefficient of the smaller matrices used for calculations will $= 0$. Thus call the top row vector $v = [x_1^2 \ x_1x_2 \ \cdots \ x_1x_k]$. The determinate calculation will then have $x_1^2(0) - x_1x_2(0) - \cdots - x_1x_k(0) = 0$.

One can also see from the above proposition and proof that v has x_1 in every term. This shows then all columns are linearly dependent on the first row of S .

3 Q5

3.1 Commonality:

They both deal with finding a hyperplane that will separate the classes correctly within some error.

3.2 Differences:

In LDA, the separating hyperplane separates the means of the samples in each class while in SVM, finds the hyperplane with the maximum of the minimum of distance between the hyperplane to the nearest point in each class.