1. Let's consider a ``***subset selection problem***'' wherein we want to do a forward search. In other words, we want to *add* one feature at a time. Assume that all features of the training data can be characterized by a multivariate Normal distribution and the discriminant is a function of the estimated covariance matrix. Show how the new $\mathbf{S}_{new}^{-1}$ can be calculated from $\mathbf{S}_{old}^{-1}$?

Answer: When we add a new feature, we are adding a row and column to the covariance matrix. It can be shown that if we partition a symmetric nonsingular matrix **A** into

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{12}^T & a_{22} \end{bmatrix}$$

the inverse is given by

$$\mathbf{A}^{-1} = \frac{1}{b} \begin{bmatrix} b\mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\boldsymbol{a}_{12}\boldsymbol{a}_{12}^T\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\boldsymbol{a}_{12} \\ -\boldsymbol{a}_{12}^T\mathbf{A}_{11}^{-1} & 1 \end{bmatrix}$$

where $b = a_{22} - \boldsymbol{a}_{12}^T\mathbf{A}_{11}^{-1}\boldsymbol{a}_{12}$. So given $\mathbf{A}_{11}$ and $\mathbf{A}_{11}^{-1}$, we can easily calculate $\mathbf{A}^{-1}$ when $\boldsymbol{a}_{12}$ and a_{22} (covariances of the new variable with the existing variables, and its variance) are added. Another possibility is to break down the Mahalanobis distance calculation (if we have the previous values stored). Let **A** be partitioned as (T. Cormen, C. Leiserson, R. Rivest, C. Stein. 2001. *Introduction to Algorithms*, The MIT Press. 2nd edition, p. 761)

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_k & \mathbf{B}^T \\ \mathbf{B} & \mathbf{C} \end{bmatrix}$$

where $\mathbf{A}_k$ is the leading $k \times k$ submatrix of $\mathbf{A}$ and $\boldsymbol{x}$ similarly broken into two parts, we can write

$$\begin{aligned}
\boldsymbol{x}^T \mathbf{A} \boldsymbol{x} &= [\boldsymbol{y}^T \boldsymbol{z}^T] \begin{bmatrix} \mathbf{A}_k & \mathbf{B}^T \\ \mathbf{B} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix} \\
&= (\boldsymbol{y} + \mathbf{A}_k^{-1} \mathbf{B}^T \boldsymbol{z})^T \mathbf{A}_k (\boldsymbol{y} + \mathbf{A}_k^{-1} \mathbf{B}^T \boldsymbol{z}) + \boldsymbol{z}^T (\mathbf{C} - \mathbf{B} \mathbf{A}_k^{-1} \mathbf{B}^T) \boldsymbol{z}
\end{aligned}$$

In our case of adding one feature, $\boldsymbol{y}$ will be d-dimensional and z will be a scalar. If we already have the d-dimensional $\boldsymbol{y}^T \mathbf{A}_k \boldsymbol{y}$ values calculated and stored, we can use the equality above to calculate the $(d+1)$-dimensional $\boldsymbol{x}^T \mathbf{A} \boldsymbol{x}$ values easily.