1 Problem 1

a)
we formulate the Lagrangian function:

$$\begin{aligned}
\mathcal{L}(\theta,\lambda) &= \sum_{k=1}^{n} ||\theta - x_k||^2 + \lambda(\theta^\intercal b - 0) = \sum_{k=1}^{n}(\theta - x_k)^\intercal(\theta - x_k) + \lambda\theta^\intercal b \\
&= \sum_{k=1}^{n}(\theta^\intercal\theta + x_k^\intercal x_k - 2\theta^\intercal x_k) + \lambda\theta^\intercal b \\
&= n\theta^\intercal\theta + \sum_{k=1}^{n} x_k^\intercal x_k - 2\sum_{k=1}^{n}\theta^\intercal x_k + \lambda\theta^\intercal b \\
\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\theta} &= 2n\theta - 2\sum_{k=1}^{n} x_k + \lambda b \stackrel{!}{=} 0 \\
\theta^* &= \frac{2\sum_{k=1}^{n} x_k + \lambda b}{2n} = \bar{x} + \frac{\lambda b}{2n}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\lambda} &= \theta^\intercal b \stackrel{!}{=} 0 \\
\left(\bar{x} + \frac{\lambda b}{2n}\right)^\intercal b &= 0 \\
\lambda &= -2n\bar{x}^\intercal b(b^\intercal b)^{-1}
\end{aligned}$$

plug λ into the formular of θ^*:

$$\theta^* = \bar{x} - \bar{x}^\intercal b(b^\intercal b)^{-1} b$$

Geometrical interpretation: The new minimizing parameter θ^* is the linear combination of the unconstrained parameter θ, which means the value of θ^* is influenced by b as well, the special case is that when b is a scaler, θ^* will equal to zero.

b)
we formulate the Lagrangian function:

$$\mathcal{L}(\theta,\lambda) = \sum_{k=1}^{n} ||\theta - x_k||^2 + \lambda(||\theta - c||^2 - 1) = \sum_{k=1}^{n}(\theta - x_k)^\intercal(\theta - x_k) + \lambda[(\theta - c)^\intercal(\theta - c) - 1]$$

$$= \sum_{k=1}^{n}(\theta^\intercal\theta + x_k^\intercal x_k - 2\theta^\intercal x_k) + \lambda(\theta^\intercal\theta + c^\intercal c - 2\theta^\intercal c - 1)$$

$$= n\theta^\intercal\theta + \sum_{k=1}^{n} x_k^\intercal x_k - 2\sum_{k=1}^{n}\theta^\intercal x_k + \lambda(\theta^\intercal\theta + c^\intercal c - 2\theta^\intercal c - 1)$$

$$\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\theta} = 2n\theta - 2\sum_{k=1}^{n} x_k + \lambda(2\theta - 2c) \overset{!}{=} 0$$

$$\theta^* = \frac{\sum_{k=1}^{n} x_k + \lambda c}{n+\lambda} = \frac{n\bar{x} + \lambda c}{n+\lambda}$$

$$\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\lambda} = ||\theta - c||^2 \overset{!}{=} 0 \qquad \text{plug } \theta^* \text{ into the euquation}$$

$$= \left|\left|\frac{n\bar{x} + \lambda c}{n+\lambda} - c\right|\right|^2 = \left|\left|\frac{n\bar{x} + \lambda c}{n+\lambda} - \frac{c(n+\lambda)}{n+\lambda}\right|\right|^2 = n^2\left|\left|\frac{\bar{x} - c}{n+\lambda}\right|\right|^2 = 1$$

$$(n+\lambda)^2 = n^2||\bar{x} - c||^2$$

$$\lambda^* = \pm n||\bar{x} - c|| - n \qquad \text{plug } \lambda^* \text{ into the euquation of } \theta^*$$

$$\theta^* = \frac{n\bar{x} + (\pm n||\bar{x} - c|| - n)c}{n \pm n||\bar{x} - c|| - n}$$

$$= \frac{\bar{x} - c \pm ||\bar{x} - c||}{\pm||\bar{x} - c||}$$

$$= \frac{\bar{x} - c}{||\bar{x} - c||} \pm c$$

Geometrical interpretation: The new θ^* is shifted by c and normalized, it's restricted to a circle with center c and radius 1.

2 Problem 2

a)
We know that if A is an $n \times n$ matrix and let λ_1,..,λ_n be its eigenvalues. Here $det(A)$ is the determinant of the matrix A and $tr(A)$ is the trace of the matrix A, and the determinant of A is the product of its eigenvalues, and the trace of A is the sum of the eigenvalues. Hence, the trace of S equals to the sum of

eigenvalues, and $S_i i$ is the diagonal elements of the scatter matrix, therefore:

$$tr(S) = \sum_{i=1}^{d} \lambda_i = \sum_{i=1}^{d} S_{ii}$$

It holds that $\lambda_i \geq 0, \forall i = 1, ..., d$, since S is positive semi-definite,

$$\lambda_1 \leq \sum_{i=1}^{d} \lambda_i = \sum_{i=1}^{d} S_{ii}$$

b)
It holds that $\lambda_1 = \sum_{i=1}^{d} S_{ii}$ if and only if $\lambda_i = 0, \forall i = 2, ..., d$. In this case the matrix S is of rank 1, which means that all features are linearly dependent.

c)
we could rewrite the matrix $\mathbf{S}$ on below,

$$\mathbf{S} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{\intercal}$$

the matrix $\mathbf{T}$ contains the eigenvectors $\tau_{\mathbf{1}}, ..., \tau_{\mathbf{d}}$ and $\mathbf{T}^{\intercal}\mathbf{T} = \mathbf{I}_d$. the matrix $\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_d)$ contains the eigenvalues, and λ_1 is the biggest eigenvalue.

For any vector ν, and $||\nu|| = 1$ we rewrite:

$$\nu^{\intercal}\mathbf{S}\nu = \nu^{\intercal}\mathbf{T}\mathbf{\Lambda}\mathbf{T}^{\intercal}\nu$$

we define $\omega = \nu^{\intercal}\mathbf{T}$,

$$\begin{aligned}
\nu^{\intercal}\mathbf{S}\nu = \nu^{\intercal}\mathbf{T}\mathbf{\Lambda}\mathbf{T}^{\intercal}\nu &= \omega^{\intercal}\mathbf{\Lambda}\omega \\
&= \lambda_1\omega_1^2 + ... + \lambda_d\omega_d^2 \\
&\leq \lambda_1(\omega_1^2 + ... + \omega_d^2) \\
&= \lambda_1||\omega||^2 = \lambda_1\omega^{\intercal}\omega \\
&= \lambda_1\nu^{\intercal}\mathbf{T}\mathbf{T}^{\intercal}\nu = \lambda_1\nu^{\intercal}\nu \\
&= \lambda_1||\nu||^2 = \lambda_1 = \lambda_{max}
\end{aligned}$$

Therefore, if we let $\nu = \tau_i$,

$$\mathbf{S_{ii}} = \tau_i^{\intercal}\mathbf{S}\tau_i \leq \lambda_1$$

for any $i = 1, ..., d$, which implies $\max_{i=1}^{d} \mathbf{S} \leq \lambda_1$

d)
The conditon is that $\mathbf{S}_{ii} = 0$, which means the all the upper right and lower left elements of $\mathbf{S}$ are zero.

3 Problem 3

a)
we formulate the formulas from question:

$$
\begin{aligned}
J(\mathbf{w}) &= ||\mathbf{Sw}|| - \frac{1}{2}\mathbf{w}^\intercal\mathbf{Sw} \\
\mathbf{v} &= \mathbf{S}^{\frac{1}{2}}\mathbf{w} \qquad \mathbf{w} = \mathbf{S}^{-\frac{1}{2}}\mathbf{v} \\
J(\mathbf{w}) &= ||\mathbf{S}\mathbf{S}^{-\frac{1}{2}}\mathbf{v}|| - \frac{1}{2}(\mathbf{S}^{-\frac{1}{2}}\mathbf{v})^\intercal\mathbf{S}\mathbf{S}^{-\frac{1}{2}}\mathbf{v} \\
&= ||\mathbf{S}^{\frac{1}{2}}\mathbf{v}|| - \frac{1}{2}\mathbf{v}^\intercal\mathbf{v} \\
&= ||\mathbf{S}^{\frac{1}{2}}\mathbf{v}|| - \frac{1}{2}||\mathbf{v}||^2 \\
\frac{\partial J(\mathbf{v})}{\partial \mathbf{v}} &= \frac{\mathbf{S}^{\frac{1}{2}}\mathbf{v}}{||\mathbf{S}^{\frac{1}{2}}\mathbf{v}||}\mathbf{S}^{\frac{1}{2}} - \mathbf{v} \\
\mathbf{v} &\leftarrow \mathbf{v} + \gamma\frac{\partial J}{\partial \mathbf{v}} = \mathbf{v} + \gamma\frac{\mathbf{S}^{\frac{1}{2}}\mathbf{v}}{||\mathbf{S}^{\frac{1}{2}}\mathbf{v}||}\mathbf{S}^{\frac{1}{2}} - \mathbf{v} \\
\mathbf{v} &\leftarrow \gamma\frac{\mathbf{Sv}}{||\mathbf{S}^{\frac{1}{2}}\mathbf{v}||} \\
&\text{If } \gamma \text{ is identity matrix} \\
\mathbf{S}^{-\frac{1}{2}}\mathbf{v} &\leftarrow \mathbf{S}^{-\frac{1}{2}}\frac{\mathbf{Sv}}{||\mathbf{S}^{\frac{1}{2}}\mathbf{v}||} = \frac{\mathbf{S}^{\frac{1}{2}}\mathbf{v}}{||\mathbf{S}^{\frac{1}{2}}\mathbf{v}||} \\
\mathbf{w} &\leftarrow \frac{\mathbf{S}^{\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}\mathbf{w}}{||\mathbf{S}^{\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}\mathbf{w}||} \\
\mathbf{w} &\leftarrow \frac{\mathbf{Sw}}{||\mathbf{Sw}||}
\end{aligned}
$$

b)

$$||\mathbf{w}|| = \left|\left|\frac{\mathbf{Sw}}{||\mathbf{Sw}||}\right|\right| = \frac{||\mathbf{Sw}||}{||\mathbf{Sw}||} = 1$$