Exercise 1

We first simplify our objective

Everywhere replace argmax by argmin!!!

$$\underset{\theta}{\operatorname{argmax}} \sum_{k=1}^{N} \|\theta - x_k\|^2 = \operatorname{argmax} \sum_k \left(\|\theta\|^2 - 2\theta^T x_k\right)$$
$$= \operatorname{argmax} \sum_k \left(\|\theta\|^2 - 2\theta^T m\right)$$
$$= \operatorname{argmax} \sum_k \|\theta - m\|^2$$
$$= \underset{\theta}{\operatorname{argmax}} \frac{1}{2}\|\theta - m\|^2$$

(a) The Lagrangian is given as:

$$L(\theta, \lambda) = \frac{1}{2}\|\theta - m\|^2 + \lambda \cdot \theta^T b$$

Look for stationary points:

$$\nabla_\theta L(\theta, \lambda) = (\theta - m + \lambda \cdot b)^T \stackrel{!}{=} \mathbb{0}^T$$

$\curvearrowright$ $$\theta - m + \lambda \cdot b = \mathbb{0} \quad (*)$$

$\curvearrowright$ Multiply both sides of the equation with b^T:

$$\underbrace{b^T\theta}_{=0} - b^T m + \lambda \cdot \underbrace{b^T \cdot b}_{=1} = \underbrace{b^T \cdot \mathbb{0}}_{=0} \implies \boxed{\lambda = b^T m} \quad (**)$$

Now substitute λ from $(**)$ in $(*)$:

$$\theta - m + b^T m \cdot b = 0 \implies \boxed{\theta = m - (b^T m) \cdot b}$$

The function $\theta \mapsto \frac{1}{2}\|\theta - m\|^2$ is strictly convex

$\curvearrowright$ Found solution is global minimum.

(b) The Lagrangian is given as:

$$L(\theta, \lambda) = \frac{1}{2}\|\theta - m\|^2 + \lambda \cdot \frac{1}{2}(\|\theta - c\|^2 - 1)$$

Look for stationary points $\nabla_\theta L \overset{!}{=} 0$:

$$\begin{aligned}\nabla_\theta L(\theta,\lambda) &= \theta - m + \lambda(\theta - c)\\ &= (1+\lambda)\theta - m - \lambda \cdot c \quad \Big|\ +c-c\\ &= (1+\lambda)\theta - m - \lambda c + c - c\\ &= (1+\lambda)(\theta - c) - (m-c) \overset{!}{=} 0\end{aligned}$$

$\rightsquigarrow \quad (1+\lambda)(\theta - c) = m - c \quad (**)$

$\rightsquigarrow$ Consider $\|\cdot\|^2$ of both sides of equation:

$$(1+\lambda)^2 \underbrace{\|\theta - c\|^2}_{=1} = \|m-c\|^2$$

$$\Rightarrow \quad 1+\lambda = \pm\|m-c\| \Rightarrow \boxed{\lambda = \pm\|m-c\| - 1} \quad (*)$$

Insert $(*)$ into $(**)$:

$$\boxed{\theta = c \pm \frac{m-c}{\|m-c\|}} \quad (***)$$

By setting both solutions $(***)$ into objective $J(\theta) = \frac{1}{2}\|\theta - m\|^2$ we observe:

$$J\left(c + \frac{m-c}{\|m-c\|}\right) \leq J\left(c - \frac{m-c}{\|m-c\|}\right)$$

Exercise 2

(a) $\frac{1}{N}\sum_{k=1}^{N}\left[u^T x_k - \frac{1}{N}\sum_{l=1}^{N} u^T x_l\right]^2$

„$(a-b)^2 = a^2 - 2ab + b^2$"

$= \frac{1}{N}\sum_k u^T x_k \cdot x_k^T u - 2u^T x_k \cdot m^T u + u^T m \cdot m^T u$

$\overset{(*)}{=} \frac{1}{N}\sum_k u^T\left[x_k \cdot x_k^T - 2x_k \cdot m^T + m \cdot m^T\right] \cdot u$

$= \frac{1}{N}\sum_k u^T\left[(x_k - m)\cdot(x_k - m)^T\right] \cdot u$

$= u^T \cdot \left[\frac{1}{N}\sum_k (x_k - m)\cdot(x_k - m)^T\right] \cdot u = u^T \cdot S \cdot u$

Note in $(*)$: $-x_k \cdot m^T - m \cdot x_k^T \neq -2x_k \cdot m^T$

but $\sum_k -x_k \cdot m^T - m \cdot x_k^T = \sum_k -2x_k \cdot m^T$

(b) $L(u, \lambda) = u^T S u - \lambda(\|u\|^2 - 1)$

$\nabla_u L(u, \lambda) = (2Su - 2\lambda u)^T \overset{!}{=} 0^T$

$\Rightarrow \boxed{Su = \lambda \cdot u}$

Furthermore:

$u^T S u = \lambda \cdot \underbrace{u^T u}_{=\|u\|^2 = 1} = \lambda$

That is, λ corresponds to the highest eigenvalue

Exercise 3

(a) Note

(I) $\lambda_i = u_i^T S u_i = \sum_k (u_i^T (x_k - m))^2 \geq 0 \quad \forall i$

(II) $\mathrm{Trace}(S) = \sum_{i=1}^{d} \lambda_i$

$\rightsquigarrow$ Together:

$$\sum_i S_{ii} = \mathrm{Trace}(S) \overset{(II)}{=} \sum_i \lambda_i \overset{(I)}{\geq} \lambda_1$$

(b) $\lambda_2 = \dots = \lambda_d = 0$

$\rightsquigarrow$ data lives on a one-dimensional subspace

(c)
$$\lambda_1 = \max_{u : \|u\| = 1} u^T S u \geq \max_{u \in \{e_1, \dots, e_d\}} u^T S u$$
$$= \max_{i=1}^{d} e_i^T S_e$$
$$= \max_{i=1}^{d} S_{ii}$$

(d) $u_1 \in \{e_1, \dots, e_d\}$

where $e_1, \dots, e_d$ are the canonical unit vectors.

Exercise 4

(a)

$$\varepsilon_k(w^{(t+1)}) = \left| \frac{w^{(t+1)^T} u_k}{w^{(t+1)^T} u_1} \right|$$

$$= \frac{\left(\frac{S w^{(t)}}{\| S w^{(t)} \|}\right)^T u_k}{\left(\frac{S w^{(t)}}{\| S w^{(t)} \|}\right)^T u_1}$$

$$= \frac{w^{(t)^T} \cdot S u_k}{w^{(t)^T} S u_1}$$

$$= \frac{w^{(t)^T} u_k \cdot \lambda_k}{w^{(t)^T} u_1 \cdot \lambda_1}$$

$$= \left| \frac{w^{(t)^T} u_k}{w^{(t)^T} u_1} \right| \cdot \left| \frac{\lambda_k}{\lambda_1} \right|$$

$$= \varepsilon_k(w^{(t)}) \cdot \left| \frac{\lambda_k}{\lambda_1} \right|$$

$\Rightarrow$

$$\varepsilon_k(w^{(T)}) = \varepsilon_k(w^{(T-1)}) \cdot \left| \frac{\lambda_k}{\lambda_1} \right| = \ldots = \varepsilon_k(w^{(0)}) \left| \frac{\lambda_k}{\lambda_1} \right|^T$$

$\square$

```python
import numpy,time,utils

def pca_svd(D):

    tini = time.time()

    # Compute SVD
    X = D.reshape(len(D),-1).T
    X -= X.mean(axis=1,keepdims=True)
    U,L,V = numpy.linalg.svd(X/X.shape[1]**.5,full_matrices=False)

    tfin = time.time()
    print('Time: %.3f seconds'%(tfin-tini))

    utils.scatterplot(U[:,0].dot(X),U[:,1].dot(X),xlabel='PCA 1',ylabel='PCA 2')
    utils.render(U[:,:60].T,15,4)
```

```python
def pca_powit(D):

    tini = time.time()

    # Preprocess data
    X = D.reshape(len(D),-1).T
    X -= X.mean(axis=1,keepdims=True)

    # Generate initial random vector
    w = numpy.random.mtrand.RandomState(346537).normal(0,1,[X.shape[0]])
    w = w / (w**2).sum()**.5

    # Compute covariance matrix
    S = numpy.dot(X,X.T) / X.shape[1]

    # Run iterative PCA
    Jold = float('nan')
    for it in range(25):

        Sw = numpy.dot(S,w)
        J = numpy.dot(w,Sw)

        print('iteration %2d   J(w) = %10.3f'%(it,J))

        if J - Jold < 0.01:
            print('stopping criterion satisfied')
            break
        else:
            w = Sw / (Sw**2).sum()**.5
            Jold = J

    tfin = time.time()

    print('Time: %.3f seconds'%(tfin-tini))

    return utils.render(w,1,1)
```