EECS 127/227AT 2/15/2018

L. El Ghaoui

Quiz 1: Solutions

$\sigma = \frac{1}{\sqrt{2}} \quad \frac{1}{3\sqrt{2}}$

$U \Lambda U^{\top}$

$U_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

1. *Ellipse.* Consider the ellipse

$$\mathcal{E} = \left\{ x \;:\; (x - x_0)^{\top} P^{-1} (x - x_0) \le 1 \right\},$$

where

$$x_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\top} + \frac{1}{3\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}^{\top}.$$

Determine

$U = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

(a) the center of the ellipse, $\hat{x}$;

(b) the semi-axes lengths, ρ_1, ρ_2;

(c) the corresponding principal directions u_1, u_2;

(d) an invertible matrix A and vector b such that in the coordinate system $\tilde{x} := Ax + b$, the ellipse looks like a sphere of radius 1 and center 0. (You need not provide the results in numerical form.)

Make sure to justify your answers carefully.

Solution:

(a) The center of the ellipse is x_0.

(b) The matrix P is symmetric and has the eigenvalue decomposition $P = U \Lambda U^{\top}$, where $U = [u_+, u_-]$, $u_{\pm} = (1/\sqrt{2})(1, -1)$, $\Lambda = \operatorname{diag}(1, 1/3)$. Hence u_+ (resp. u_-) is a normalized eigenvector corresponding to the eigenvalue 1 (resp. 3). Defining $\bar{x} := U^{\top}(x - x_0)$, the condition $x \in \mathcal{E}$ translates as

$$(x - x_0)^{\top} P^{-1} (x - x_0) = (x - x_0)^{\top} U \Lambda^{-1} U^{\top} (x - x_0) = \bar{x}_1^2 + 3\bar{x}_2^2 \le 1. \tag{1}$$

From this, we determine the semi-axis lengths to be $\rho_1 = 1$, $\rho_2 = 1/\sqrt{3}$.

(c) From the above we see that the principal directions correspond to setting $\bar{x} = e_1 = (1, 0)$ and $\bar{x} = e_2 = (0, 1)$ respectively. In x-space, this corresponds to setting $x - x_0 = U e_i$, $i = 1, 2$. With $U e_1 = u_+$, $U e_2 = u_-$, we obtain that u_+, u_- are the principal directions.

(d) We set $\tilde{x} = \Lambda^{-1/2} \bar{x}$, so that the equation (1) becomes $\|\tilde{x}\|_2^2 \le 1$. Thus

$$\tilde{x} = \Lambda^{-1/2} U^{\top} (x - x_0) = Ax + b,$$

where

$$A = \Lambda^{1/2} U^{\top}, \quad b = -\Lambda^{1/2} U^{\top} x_0.$$

2. *Optimal weighting in a test.* A $n \times m$ matrix M contains the scores of n students on a test having m parts, so that M_{ij} is the score of student i in part j. We define a vector $w \in \mathbb{R}^m$ containing the weights w_j associated with each part $j = 1, \ldots, m$.

   (a) Express the vector $s \in \mathbb{R}^n$ containing the score of each student, in terms of M and w, in matrix notation.

   (b) Express the vector $\hat{m}$ containing the scores obtained in each part on average across students. You may denote by $m_i \in \mathbb{R}^m$ the i-th column of $M^\top$, $i = 1, \ldots, n$.

   (c) Someone suggests to the professor to use the maximum variance principle in order to compute a weight vector w. Explain how to do so, making sure to detail the covariance matrix involved.

   (d) What are possible shortcomings of the maximum-variance approach? *Hint:* comment on the sign of the entries of the maximum-variance weight vector w.

Solution:

(a) We have $s = Mw$.

(b) We have $\hat{m} = M^\top e$, where $e = (1/n)\mathbf{1}$.

(c) The variance of the scores is

avg. score of all students.

$$\sigma(w) = \frac{1}{n}\sum_{i=1}^{n}(s_i - \hat{s})^2,$$

where $\hat{s} = (1/n)\mathbf{1}^\top s$. We have

$$\sigma(w) = w^\top C w,$$

where C is the $n \times n$ symmetric matrix

$$C = \frac{1}{n}\sum_{i=1}^{n}(m_i - \hat{m})(m_i - \hat{m})^\top,$$

with $m_i^\top$ the i-th row of M, and

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n} m_i = (1/n)M^\top \mathbf{1}.$$

We then solve

$$\max_w \; w^\top C w \; : \; \|w\|_2 = 1,$$

and set w to the eigenvector corresponding to the largest eigenvalue.

(d) It may turn out that the maximum-variance vector has some negative component, which would not make sense as a weighting vector.

3. *PCA and optimal projection on a line.* In this exercise, we show the equivalence between PCA and a kind of least-squares problem involving a line.

We consider a matrix of data points $X = [x_1, \ldots, x_m] \in \mathbb{R}^{n,m}$, and seek to find a line such that the sum of squared distances from the points to the line is minimized. In the sequel, we parametrize a generic line in $\mathbb{R}^n$ as

$$\mathcal{L}(x_0, u) = \{x_0 + tu \ : \ t \in \mathbb{R}^n\},$$

where $x_0, u \in \mathbb{R}^n$ are given, with $\|u\|_2 = 1$. Geometrically, u provides the direction of the line, and x_0 its intercept.

(a) Show that the distance from a given line $\mathcal{L}(x_0, u)$ to a given point $x \in \mathbb{R}^n$ is given by

$$D(x, \mathcal{L}(x_0, u))^2 = (x - x_0)^\top P(u)(x - x_0),$$

where $P(u) := I_n - uu^\top$.

(b) Is the symmetric matrix $P(u)$ positive semi-definite, definite? What are its eigenvalues?

(c) What is the geometric interpretation of the linear map $x \to P(u)x$?

(d) Now consider the minimization problem referred to above. Show that an optimal point x_0 is given by the center (average) of all data points. *Hint:* fix u and solve for x_0.

(e) Show that an optimal direction u is given by the standard variance maximization problem at the heart of principal component analysis:

$$\max_u \ u^\top C u \ : \ \|u\|_2 = 1,$$

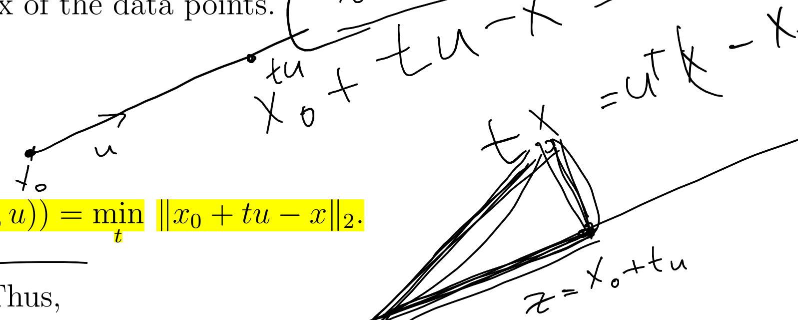
where C is the covariance matrix of the data points.

$x_0 + tu$ $= L$

$x_0 + tu - x =$ $= u^\top(x - x_0)$

Solution:

(a) The distance is given by



$$D(x, \mathcal{L}(x_0, u)) = \min_t \ \|x_0 + tu - x\|_2.$$

At optimum, $t^* = u^\top(x - x_0)$. Thus,

$$D(x, \mathcal{L}(x_0, u)) = \|x_0 - x + u(u^\top(x - x_0))\|_2 = \|P(u)(x - x_0)\|_2.$$

Exploiting the fact that $P(u)^2 = P(u)$ leads to the desired result.

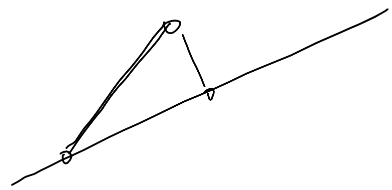
(b) Since $P(u)$ is symmetric, and satisfies $P(u)^2 = P(u)$, its eigenvalues are either 0 or 1. Hence it is PSD, but not positive definite, since $P(u)z = 0$ whenever $z^\top u = 0$.

(c) For a n-vector x,

$$P(u)x = x - (x^\top u)u.$$

Here $(x^\top u)$ is the component of x along direction u, and $(x^\top u)u$ is the projection of x on the line with direction u passing through 0, $\mathcal{L}(0, u)$. This $P(u)x$ is difference between x and its projection on that line.

(d) The optimization problem reads



$$\begin{aligned} &\min_{u\,:\,\|u\|_2=1,\, x_0} \sum_{i=1}^m D(x, \mathcal{L}(x_0, u))^2 \\ = &\min_{u\,:\,\|u\|_2=1,\, x_0} \sum_{i=1}^m (x_i - x_0)^\top P(u)(x_i - x_0). \end{aligned}$$

For a fixed u, the problem of minimizing the above in terms of x_0 is unconstrained, convex and differentiable. Zeroing the gradient characterizes optimal points:

$$\sum_{i=1}^m P(u)(x_i - x_0) = m \cdot P(u)(\hat{x} - x_0) = 0,$$

where

$$\hat{x} := \frac{1}{m} \sum_{i=1}^m x_i$$

is the average point. We observe that $x_0 = \hat{x}$ is optimal. Note that the minimizer is not unique; only $\hat{x}$ works for any u.

(e) We have

$$\sum_{i=1}^m (x_i - \hat{x})^\top P(u)(x_i - \hat{x}) = -\sum_{i=1}^m \left(u^\top (x_i - \hat{x})\right)^2 + \text{cst}.$$

This leads to the desired result.