Basic algebra for data analysis

Marco Corneli

Exercise 1. Consider a symmetric matrix $S \in \mathbb{R}^{D \times D}$ and the associated quadratic form $f : \mathbb{R}^D \to \mathbb{R}$ defined by

$$f(u) := u^T S u = \sum_{i=1}^{D} \sum_{j=1}^{D} u_i u_j S_{ij}$$

(a) Compute the partial derivative $\frac{\partial f}{\partial u_{i_0}}(u)$, for a generic $i_0 \in \{1, \ldots, D\}$ and use it to conclude that

$$\nabla f(u) = 2Su \tag{1}$$

(b) Use the above point to show that, given $g(u) := u^T u$, then $\nabla g(u) = 2u$.

It is now also assumed that S is positive definite and its eigen-decomposition is $S = Q \Lambda Q^T$, where Q is the orthogonal matrix of the eigenvectors and Λ has on the main diagonal the eigenvaules, assumed to be sorted in decreasing order $(\lambda_1 > \lambda_2 > \ldots, \lambda_D)$. We introduce the set

$$\mathcal{U} := \{u \in \mathbb{R}^D | \parallel u \parallel_2 = 1\}.$$

(a) Show that if $u \in \mathcal{U}$ then $y := Q^T u$ is still in $\mathcal{U}$.

(b) Use the above point to prove that

$$\max_{u \in \mathcal{U}} (u^T S u) \leq \lambda_1. \tag{2}$$

Exercise 2. Consider a vector space V of dimension N and sub-vector space $W \subset V$ provided with an orthonormal basis $\{u_1, \dots, u_K\}$. Recalling the definition of orthogonal projection we gave in class, $P_W : V \mapsto W$

$$P_W(v) = \sum_{j=1}^{K} \langle v, u_j \rangle u_j,$$

show that

(a) the projection matrix associated to P_W can be written in the form $M_{P_W} = UU^T$. Who are the columns of U?

(b) Prove that M_{P_W} is symmetric and idempotent[1].

Exercise 3. [*Principal Component Analysis.*] Given N observations $x_1, \dots, x_N$ in $\mathbb{R}^D$, we introduce the empirical mean

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad \in \mathbb{R}^D$$

and the empirical variance (matrix)

$$\overline{S}_x = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})(x_i - \overline{x})^T, \qquad \in \mathbb{R}^{D \times D}$$

which is symmetric and positive definite. The aim of the exercise is to detect a matrix $U \in \mathbb{R}^{D \times K}$ (with $K \ll D$), with the same properties of the one in Exercise 2-(a) (which ones?), inducing the transformation $y_i = U^T x_i \in \mathbb{R}^K$ and maximizing the variance of the the low dimensional data $y_1, \dots, y_N$.

(a) Prove that $\overline{y} = U^T \overline{x}$.

(b) Prove that $\overline{S}_y = U^T \overline{S}_x U$.

We now want our U to maximize trace$(\overline{S}_y)$.

[1] Namely that $M_{P_W} M_{P_W} = M_{P_W}$.

(a) Motivate why the maximization problem can be stated as

$$\max_{\substack{u_1,\ldots,u_K \\ u_i \perp u_j}} \sum_{k=1}^{K} \left(u_k^T S_x u_k - \lambda_k \parallel u_k \parallel_2^2 \right), \qquad \lambda_k \in \mathbb{R}.$$

Who are $u_1, \ldots, u_K$?

(b) Assume first that $K = 1$ and use Eq. (1) to find stationary points for the objective function $g(u_1, \lambda_1) := u_1^T S_x u_1 - \lambda_1 \parallel u_1 \parallel_2^2$.

(c) Use Eq. (2) to choose the stationary point maximizing g.

(d) Deduce the solution $u_1, \ldots, u_K$ in the general case $K > 1$.