Maths for Intelligent Systems
Exercise 2

Marc Toussaint, Andrea Baisero
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

November 20, 2018

1 Projections

(a) In $\mathbb{R}^n$, a plane (through the origin) is typically described by the linear equation

$$c^\top x = 0 \ , \tag{1}$$

where $c \in \mathbb{R}^n$ parameterizes the plane. Provide the matrix that describes the orthogonal projection onto this plane. (Tip: The SVD describes matrices as sum of rank-1 matrices; here, think of the projection as **I** minus a rank-1 matrix.)

(b) In $\mathbb{R}^n$, let's have k linearly independent $\{v_i\}_{i=1}^k$, which form the matrix $V = (x_1, .., x_k) \in \mathbb{R}^{n \times k}$. Let's formulate a projection using an optimality principle, namely,

$$\alpha^*(x) = \operatorname*{argmin}_{\alpha \in \mathbb{R}^k} \|x - V\alpha\|^2 \ . \tag{2}$$

Note that $V\alpha = \sum_{i=1}^k \alpha_i v_i$ is just the linear combination of v's with coefficients α. The projection of a vector x is then

$$x_\parallel = V\alpha^*(x) \ . \tag{3}$$

Derive the equation for the optimal $\alpha^*(x)$ from the optimality principle.

2 SVD

Consider the matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \ , \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \tag{4}$$

(a) Describe their 4 fundamental spaces (dimentionality, possible basis vectors).

(b) Find the SVD of A (using pen and paper only!)

(c) Given an arbitrary input vector $x \in \mathbb{R}^3$, write the linear transformations P_A and P_B which extract its "input null space" component for each respective matrix.

(d) Compute the pseudo inverse $A^\dagger$.

(e) Determine all solutions to the linear equations $Ax = y$ and $Bx = y$ with $y = (2, 3, 0, 0)$. What is the more general expression for an arbitrary y?

3 Covariance and PCA

Suppose we're given a collection of zero-centered data points $D = \{x_i\}_{i=1}^N$, with each $x_i \in \mathbb{R}^n$. The covariance matrix is defined as

$$C = \frac{1}{n} \sum_{i=1}^{N} x_i x_i^\top = \frac{1}{n} X^\top X$$

where (consistent to ML lecture convention) the data matrix $X = (x_1^\top; ..; x_N^\top)$ contains each $x_i^\top$ as rows, $X^\top = (x_1, .., x_N)$.

If we project D onto some unit vector $v \in \mathbb{R}^n$, then the variance of the projected data points is $v^\top C v$. Show that the direction that maximizes this variance is the largest eigenvector of C. (Hint: Expand v n terms of the eigenvector basis of C and exploit the constraint $v^\top v = 1$.)

(This optimization is the first step of the so-called Principal Components Analysis (PCA).)

4 Bonus: Scalar product and Orthogonality

(a) Show that $f(x, y) = 2x_1y_1 - x_1y_2 - x_2y_1 + 5x_2y_2$ is an scalar product on $\mathbb{R}^2$.

(b) In the space of functions with the scalar product $\langle fg \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx$, what is the projection of $\sin(x)$ onto $\sin(2x)$? (Graphical argument is ok)

(c) What property does a matrix M has to satisfy in order to be a valid metric tensor, i.e. such that $x^\top M y$ is a valid scalar product?