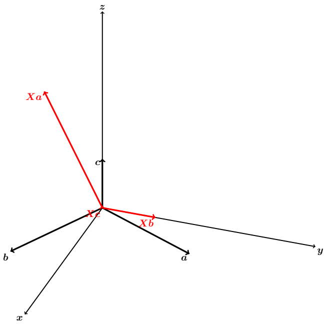
Internet Analytics (COM-308)

Homework Set 4

Exercise 1

In this exercise we study the geometric interpretation of singular value decomposition (SVD) seen in class.Assume the images of three vectors $a = [1, 2, 0]^\intercal, b = [2, -1, 0]^\intercal$ and $c = [0, 0, 1]^\intercal$ under the linear transformation $X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix}$ is $Xa = [3, 0, 4]^\intercal, Xb = [0, 1, 0]^\intercal$ and $Xc = [0, 0, 0]^\intercal$, respectively.



(a) Find the linear transformation X (matrix X) directly by solving the system of linear equations

$$X[a\ b\ c] = [Xa\ Xb\ Xc].$$

(b) The SVD decomposition of matrix X is

$$X = U\Sigma V^\intercal,$$

where we can interpret the three matrices U, Σ and $V^\intercal$ as follows:(i) the two matrices U and $V^\intercal$ are changes of basis; and (ii) matrix Σ is scaling each axis (dimension) independently. Determine the transformation X from this interpretation of SVD (without any calculations).

Exercise 2

Let X be an $n \times m$ matrix, where each of the n rows is a datum represented by an m-dimensional vector. Furthermore, assume that X is zero-mean, i.e. each column of X sums up to zero. Consider the singular value decomposition (SVD) of X:

$$X = U\Sigma V^\intercal.$$

(a) Express the SVD of the matrix $X^\intercal$ in terms of the SVD of X.

(b) We define the $m \times m$ covariance matrix $\mathrm{Cov}_{m\times m}[X]$ as

$$\mathrm{Cov}_{m\times m}[X] = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \\ \mathrm{Cov}(X_2, X_1) & \ddots & & \\ \vdots & & & \mathrm{Cov}(X_m, X_m) \end{pmatrix}$$

where $\mathrm{Cov}(X_i, X_j) = \frac{1}{n}\sum_{k=1}^{n} X_{ki}X_{kj}$. Express $\mathrm{Cov}_{m\times m}[X]$ in terms of multiplication of two matrices.

(c) Principal Component analysis (PCA) can be seen as the eigendecomposition of the covariance matrix:

$$\mathrm{Cov}_{m\times m}[X] = Q\Lambda Q^{\intercal}.$$

How does this interpretation of PCA relate to the SVD of X? Express Q and Λ in terms of U, Σ and V.

(d) With PCA, we project the original data X into a new space $Y = XV$ of dimension d. Compute the covariance matrix of Y.