MAT-INF4310: Mandatory assignment #3, autumn 2017

To be handed in by October 26., 14:30

You must hand in commented scripts which actually compile and work. You must also use "Devilry".

We shall consider the following problem:
Given n points in $\mathbb{R}^m$; $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, and an integer $k \leq m$, find a k dimensional subspace $W \subset \mathbb{R}^m$ such that

$$\sum_{j=1}^{n} \operatorname{dist}(\mathbf{x}_j - W)^2$$

is minimal.

To fix the notation set $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T$ for $j = 1, \dots, n$, and define the matrix $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ such that the ijth entry of $\mathbf{X}$ is x_{ij}.

Assume first that W is any k-dimensional subspace of $\mathbb{R}^m$, let $\{\mathbf{w}_j\}_{j=1}^{k}$ be an orthonormal basis of W and we can use Gram-Schmidt to find $\{\mathbf{w}_j\}_{j=k+1}^{m}$ such that $\{\mathbf{w}_j\}_{j=1}^{m}$ is an orthonormal basis for $\mathbb{R}^m$. In this way we can write

$$\mathbb{R}^m = W \oplus W^{\perp}.$$

For $\mathbf{x} = \mathbb{R}^m$ we let $\operatorname{proj}_{W^\perp}(\mathbf{x})$ denote the orthogonal projection onto $W^\perp$.

Exercise 1. Show that

$$\sum_{i=1}^{n} \|\operatorname{proj}_{W^\perp}(\mathbf{x}_i)\|_2^2 = \left\|\mathbf{X}^T\mathbf{W}\right\|_F^2,$$

with $\|\cdot\|_F$ denoting the Frobenius norm and (beware of bold face notation) $\mathbf{W}$ is the $m \times (m-k)$ matrix $\mathbf{W} = [\mathbf{w}_{k+1}\ \mathbf{w}_{k+2}\ \dots\ \mathbf{w}_m]$.

Therefore $\left\|\mathbf{X}^T\mathbf{W}\right\|_F$ measures "how well" the linear subspace W fits the "observations" X.

Exercise 2.

a) Show that

$$\left\|\mathbf{X}^T\mathbf{W}\right\|_F^2 = \left\|\Sigma^T\mathbf{U}^T\mathbf{W}\right\|_F^2,$$

where $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$ is the singular value decomposition of $\mathbf{X}$.

b) Let the ith column of $\mathbf{U}$ be $\mathbf{u}_i$ so that $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_m]$. Show that

$$\left\|\mathbf{X}^T\mathbf{W}\right\|_F^2 = \sum_{i=1}^{m} \sigma_i^2 \|\operatorname{proj}_{W^\perp}(\mathbf{u}_i)\|_2^2,$$

where σ_i, $i = 1, \dots, m$ are the singular values of $\mathbf{X}$. We use the convention that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$.

c) Explain why

$$m - k = \|\mathbf{W}\|_F^2 = \sum_{i=1}^{m} \|\operatorname{proj}_{W^\perp}(\mathbf{u}_i)\|^2.$$

Exercise 3.

a) Explain how finding the "distance to the best subspace"; $\left\|\mathbf{X}^T\mathbf{W}\right\|_F$, can be accomplished by solving the constrained minimization problem:

$$\text{Minimize } \sum_{i=1}^{m} \sigma_i^2 \kappa_i, \tag{1}$$
$$\text{subject to the conditions } 0 \le \kappa_i \le 1,\ i = 1, \dots, m \text{ and } \sum_{i=1}^{m} \kappa_i = m - k.$$

b) Explain why the solution to the minimization problem (1) is

$$\kappa_1 = \kappa_2 = \cdots = \kappa_k = 0, \quad \kappa_{k+1} = \cdots = \kappa_m = 1.$$

c) How can we choose W so that $\left\|\mathbf{X}^T\mathbf{W}\right\|_F$ is minimized?

d) Implement a routine "`[B,e]=bestfit(X,k)`", which given the matrix $\mathbf{X}$ and an integer $k \le m$, computes a matrix $\mathbf{B}$ whose columns span the "best" k-dimensional subspace W, and the distance $\left\|\mathbf{X}^T\mathbf{W}\right\|_F$ (You may use a ready-made routine for the singular value decomposition.)

Exercise 4. We can also try to find a fit to the "observations" $\{\mathbf{x}_j\}_{j=1}^n$ by "vanilla least squares". In this case we view the the first k components of $\mathbf{x}$ as input data, and the last $m - k$ components as output data. Set $\mathbf{Y} = \mathbf{X}_{1:k,1:n}$ and $\mathbf{Z} = \mathbf{X}_{k+1:m,1:n}$. Then $\left\{(\mathbf{y}, \mathbf{A}\mathbf{y}) \mid \mathbf{y} \in \mathbb{R}^k\right\}$ where $\mathbf{A} \in \mathbb{R}^{(m-k)\times k}$, is a k-dimensional subspace of $\mathbb{R}^m$. Vanilla least squares means choosing the matrix $\mathbf{A}$ so that

$$\|\mathbf{A}\mathbf{Y} - \mathbf{Z}\|_F \quad \text{is minimized.}$$

a) Show that the solution to this satisfies

$$\mathbf{A}^T = (\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}\mathbf{Z}^T.$$

b) Implement a routine "`[A,e]=vanleastsqr(X,k)`" which given the matrix $\mathbf{X}$ and an integer $k < m$, computes the matrix $\mathbf{A}$ and the error $\|\mathbf{A}\mathbf{Y} - \mathbf{Z}\|_F$.

Exercise 5. Compare your routines `bestfit` and `vanleastsqr` in the following two examples.

- **a)** $\mathbf{X}_{1,:} = (-1.0, -0.9, -0.8, -0.7 \dots, 0.9, 1.0)$, $\mathbf{X}_{2,:}$ consists of 21 random numbers in $[-1, 1]$ with mean zero.
- **b)** $\mathbf{X}_{1,:}$ and $\mathbf{X}_{2,:}$ are both 21 random numbers in $[-1, 1]$ with mean zero.

In both examples $k = 1$. For each example, plot the "best" linear subspaces and the points $\mathbf{X}_{:,j}$.