Linear Algebra II
Best-Fit Subspaces

Ming-Hsuan Kang

Best-fit Subspaces

Given a $m \times n$ real matrix A, let $\{\vec{x}_1^t, \cdots, \vec{x}_m^t\}$ be the rows of A, which is a subset of $\mathbb{R}^n$. A k-dimensional subspace W in $\mathbb{R}^n$ is called a best-fit k-subspace of $\{\vec{x}_1, \cdots, \vec{x}_m\}$. if

$$W \in \underset{W' : dim W' = k}{\arg\min} \left\{ \sum_{i=1}^{m} \|\vec{x}_i - \mathrm{proj}_{W'}(\vec{x}_i)\|^2 \right\}.$$

Together with the fact

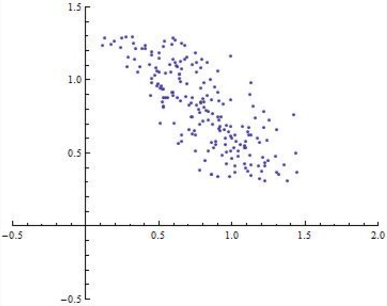
$$\|\vec{x}_i\|^2 = \|\mathrm{proj}_{W'}(\vec{x}_i)\|^2 + \|\vec{x}_i - \mathrm{proj}_{W'}(\vec{x}_i)\|^2,$$

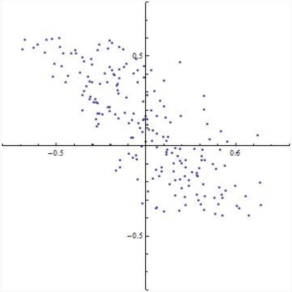
it is equivalent to

$$W \in \underset{W' : dim W' = k}{\arg\max} \left\{ \sum_{i=1}^{m} \|\mathrm{proj}_{W'}(\vec{x}_i)\|^2 \right\}.$$

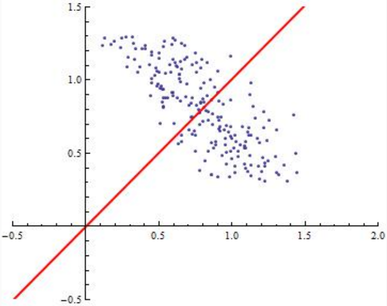
Example

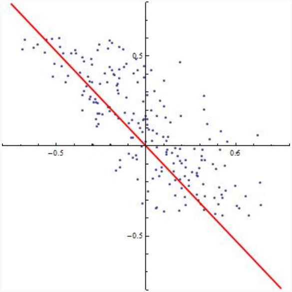
What will be the best-fit 1-subspaces for the following data points?





The following are the best-fit 1-subspaces.





Fitting Energy Function

For convenience, define the fitting energy function to be

$$E_A(W) := \sum_{i=1}^{m} \|\mathrm{proj}_W(\vec{x}_i)\|^2.$$

Note that if $\{\vec{w}_1, \cdots, \vec{w}_k\}$ is an orthonormal basis of W, then

$$E_A(W) = \sum_{i=1}^{m} \sum_{j=1}^{k} \|\mathrm{proj}_{\vec{w}_j}(\vec{x}_i)\|^2.$$

On the other hand,

$$A\vec{w}_j = \begin{pmatrix} \vec{x}_1^t \\ \vdots \\ \vec{x}_m^t \end{pmatrix} \vec{w}_j = \begin{pmatrix} \vec{x}_1^t \vec{w}_j \\ \vdots \\ \vec{x}_m^t \vec{w}_j \end{pmatrix} \quad \text{and} \quad |\vec{x}_i^t \vec{w}_j| = \|\mathrm{proj}_{\vec{w}_j}(\vec{x}_i)\| \ , \forall i.$$

Thus,

$$E_A(W) = \sum_{j=1}^{k} \|A\vec{w}_j\|^2.$$

Now we can rewrite the definition of best-fit k-subspaces.

Definition

A k-dimensional subspace W in $\mathbb{R}^n$ is called a best-fit k-subspace of the row vectors of an $m \times n$ matrix A, if

$$E_A(W) = \sum_{j=1}^{k} \|A\vec{w}_j\|^2$$

is maximal among all k-dimensional subspaces. Here $\{\vec{w}_1, \cdots, \vec{w}_k\}$ is an orthonormal basis of W.

Singular Values and Left Singular Vectors

Note that

$$\|A\vec{u}\|^2 = \vec{u}^t A^t A\vec{u}$$

and we have shown that $A^t A$ is always an $n \times n$ positive semi-definite matrix. Let

$$\lambda_1 \geq \cdots \geq \lambda_n \geq 0$$

be the list of all eigenvalues of A and $\{\vec{v}_1, \cdots, \vec{v}_n\}$ be a corresponding orthonormal eigenbasis.

Definition

The value λ_i is called the *i-th singular value* and $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is called a set of *left singular vectors*.

Remark. The set of singular values is unique but the set of left singular vectors is not unique.

Theorem

For $1 \leq i \leq n$, we have

$$\lambda_i = \max_{\vec{v}}\{\|A\vec{v}\|^2\} \quad \textit{and} \quad \vec{v}_i \in \arg\max_{\vec{v}}\{\|A\vec{v}\|^2\},$$

where $\vec{v}$ runs through all unit vectors orthogonal to $\vec{v}_j$ for all $j < i$.

Proof. Let $\vec{v}$ be a unit vector orthogonal to $\vec{v}_j$ for all $j < i$. We can express $\vec{v}$ as $a_i\vec{v}_i + \cdots + a_n\vec{v}_n$ for some a_i. Then $1 = \|\vec{v}\|^2 = a_i^2 + \cdots + a_n^2$ and

$$\begin{aligned}\|A\vec{v}\|^2 &= \vec{v}^t A^t A\vec{v} = (a_i\vec{v}_i + \cdots + a_n\vec{v}_n)^t(\lambda_i a_i\vec{v}_i + \cdots + \lambda_n a_n\vec{v}_n) \\ &= \lambda_i a_i^2 + \cdots + \lambda_n a_n^2 \leq \lambda_i(a_i^2 + \cdots + a_n^2) = \lambda_i.\end{aligned}$$

Therefore, when $\vec{v} = \vec{v}_i$, $\lambda_i = \|A\vec{v}_i\|^2$ is the maximum.

Singular Vectors and Best-fit Subspaces

Theorem

The subspace spanned by the first k left singular vectors of A is a best-fit k-subspace of the rows of A.

Proof

Let V_k be the subspace spanned by the first k left singular vectors. We shall prove the following statement by induction on k. For any k-dimensional subspace W, there exists an orthonormal basis $\{\vec{w}_1, \cdots, \vec{w}_k\}$ such that

$$\|A\vec{w}_i\| \leq \|A\vec{v}_i\| \quad \text{for all } i = 1, \cdots, k.$$

Then

$$E_A(W) = \|A\vec{w}_1\|^2 + \cdots + \|A\vec{w}_k\|^2 \leq \|A\vec{v}_1\|^2 + \cdots + \|A\vec{v}_k\|^2 = E_A(V_k).$$

Thus, V_k is a best-fit k-subspace.

For $k = 1$, by the previous theorem, $\|A\vec{v}_1\| \geq \|A\vec{w}\|$ for any unit vector $\vec{w}$. Especially, $\|A\vec{v}_1\| \geq \|A\vec{w}_1\|$ for any unit vector $\vec{w}_1$ in W.

Proof

Suppose the statement holds for all subspaces of dimensional less then k. Let W be a k-dimensional subspace. Then

$$\dim W + \dim(V_{k-1})^{\perp} = k + (n - k + 1) = n + 1 > n,$$

so $W \cap (V_{k-1})^{\perp}$ is a non-zero subspace. Choose a unit vector $\vec{w}_k$ in $W \cap (V_{k-1})^{\perp}$. Then by the previous theorem, we have

$$\|A\vec{v}_k\| \geq \|A\vec{w}_k\|.$$

Let W_0 be the orthogonal complement of $\vec{w}_k$ in W, which is of dimension $k-1$. By induction, W_0 admits an orthonormal basis $\{\vec{w}_1, \cdots \vec{w}_{k-1}\}$ such that

$$\|A\vec{v}_i\| \geq \|A\vec{w}_i\| \quad \text{for all } i = 1, \cdots, k-1.$$

Combining the above result, $\{\vec{w}_1, \cdots \vec{w}_k\}$ is the desired orthonormal basis of W. The proof is complete.