1 Fitting linear manifolds

Def 1.1 A subset $L$ of $\mathbb{R}^p$ is called a k-dimensional linear manifold (or a k-dimensional affine subspace) if there is a k-dimensional linear subspace $U$ with the following properties:

(1) $x, y \in L \Rightarrow x - y \in U$.
(2) $x \in L, u \in U \Rightarrow x + u \in L$.

If $u_1, \ldots, u_k$ form a basis of $U$, and $u_0 \in L$, then every $x \in L$ has a unique representation

$$x = u_0 + \sum_{i=1}^{k} x_i u_i$$

Without loss of generality we can assume that $u_0, \ldots, u_k$ are orthonormal.

Prop 1.1 Let $L$ denote a k-dimensional linear manifold with associated subspace $U$. Then for every $x \in \mathbb{R}^p$ there is a unique closest point $z \in L$.

Proof: For $z = u_0 + \sum z_i u_i$, we have

$$\|x - z\|^2 = \|x - u_0 - \sum z_i u_i\|^2$$

$$= \|x - u_0\|^2 + \sum z_i^2 - 2 \sum z_i \langle x, u_i \rangle$$

Taking derivatives with respect to $z_i$ gives $z_i = \langle x, u_i \rangle, i = 1, \ldots, k$.

The closest point in $L$ to $x$ thus is $z = u_0 + \sum \langle x, u_i \rangle u_i$. The squared distance of $x$ from $L$ is

$$d^2(x, L) = \|x - u_0\|^2 - \sum \langle x, u_i \rangle^2$$

Note: $x - z \in U^\perp$ because

$$\langle x - z, u_i \rangle = \langle x - u_0 - \sum \langle x, u_j \rangle u_j, u_i \rangle = 0$$

Define $V = U^\perp$, the orthogonal complement to $U$ in $\mathbb{R}^p$. Let $P_V$ denote the orthogonal projection onto $V$.

Note: $d^2(x, z) = \|x - z\|^2 = \|P_V(x - z)\|^2 = \|P_V(x) - u_0\|^2$.

The following proposition is the main result of this section:
Prop 1.2 Given a collection $x_1, \ldots, x_n$ of points in $\mathbb{R}^p$. A $k$-dimensional linear manifold $L$ minimizing $\sum_{i=1}^n d^2(x_i, L)$ has the following properties:

1) $\bar{x} = 1/n \sum x_i \in L$.

2) The associated subspace $U$ is spanned by $k$ largest eigenvectors (eigenvectors with the largest eigenvalues) of the sample covariance matrix $\Sigma = 1/n \sum (x_i - \bar{x})(x_i - \bar{x})^T$.

Note: Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ denote the eigenvalues of $\Sigma$. $L$ will be uniquely determined only if $\lambda_k > \lambda_{k+1}$. It is easy to think of point configurations in the plane for which there is no unique closest line, for example if $x_1, \ldots, x_n$ are all the same or if they are arranged in a symmetric pattern, like the vertices of a regular hexagon.

To prove Prop. 1.2, we first show that, for any fixed subspace $U$, the linear manifold with associated subspace $U$ that is closest to $x_1, \ldots, x_n$ has to go through $\bar{x}$.

Prop 1.3 Let $U$ be a linear subspace and $L$ be a linear manifold with associated subspace $U$ minimizing $\sum d^2(x_i, L)$. Then $\bar{x} \in L$.

Proof: $\sum d^2(x_i, L) = \sum \|P_U(x_i) - u_0\|^2$. Thus $u_0 = 1/n \sum P_U(x_i)$. This shows that $\bar{x} \in L$ because

$$\bar{x} = P_U(\bar{x}) + P_U(\bar{x}) = P_U(\bar{x}) + u_0.$$

Without loss of generality we can assume that $\bar{x} = 0$. We have shown that the closest linear manifold then passes through 0, i.e. is a linear subspace. We thus have reduced the problem to finding the closest linear subspace to a set of points.

Prop 1.4 A $k$-dimensional linear subspace $U$ closest to $x_1, \ldots, x_n$ is spanned by $k$ largest eigenvectors (eigenvectors with the largest eigenvalues) of $\Psi = \sum x_ix_i^T$.

Proof: Let $u_1, \ldots, u_k$ denote an orthonormal basis for $U$. Then

$$\sum d^2(x_i, U) = \sum \|x_i\|^2 - \sum_{i=1}^n \sum_{j=1}^k (x_i, u_j)^2.$$
Thus we want to maximize

$$ \sum_{i} \sum_{k} \langle x_i, u_j \rangle^2 = \sum_{j} u_j^T \sum_{i} x_i x_i^T u_j = \sum_{j=1}^{k} u_j^T \Psi u_j. $$

Let $a_1, \ldots, a_p$ denote eigenvectors of $\Psi$ for eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. We now switch to eigencoordinates. In this coordinate system, the quadratic form defined by $\Psi$ is diagonal. In other words, we want to find orthonormal vectors $v_1, \ldots, v_k$ maximizing

$$ \sum_{i=1}^{k} v_i^T \text{diag}(\lambda_1, \ldots, \lambda_p) v_i = \sum_{i=1}^{k} \sum_{j=1}^{p} v_i^2 \lambda_j, $$

Define $h_j = \sum_{i=1}^{k} v_i^2, j = 1, \ldots, p$. Obviously, $\sum_{j} h_j = k$, because $\sum_{j} h_j$ is the sum of squared norms of $k$ orthonormal vectors. Also, $0 \leq h_j \leq 1$. To see this, consider a matrix $V$ with orthonormal rows $v_1^T, \ldots, v_k^T$. $V$ can be expanded to a $p \times p$ matrix $V^*$ with orthonormal rows. $V^*$ also has orthonormal columns, and the result follows.

Now consider the function $\phi(h_1, \ldots, h_k) = \sum_{i=1}^{p} h_i \lambda_i$. Under the constraints that $0 \leq h_j \leq 1$ and $\sum_{j} h_j = k$, $\phi(h_1, \ldots, h_k)$ is maximized for $h_1, \ldots, h_k = 1$ and $h_{k+1}, \ldots, h_p = 0$. This implies that $v_1, \ldots, v_k$ lie in a space spanned by $k$ largest eigenvectors of $\Psi$.

We have so far defined a linear manifold $L$ by its associated subspace $U$ and a translation vector $u_0$:

$$ L = \{u_0 + u : u \in U\}, $$

or, for $u_1, \ldots, u_k$ a basis of $U$,

$$ L = \{u_0 + \sum x_i u_i\}. $$

The mapping $\phi : \mathbb{R}^k \to \mathbb{R}^p$ defined by

$$ \phi(x) = u_0 + \sum_{i=1}^{k} x_i u_i $$
is a homeomorphism between \( L \) and \( R^k \). A linear manifold thus has a global chart (global parametrization).

We will now discuss an alternative representation for linear manifolds, as the kernel of an affine map.

**Prop 1.5** Let \( l : R^p \to R^q \) with \( q < p \) denote a linear map of full rank. Then for any \( c \in R^q \) the set

\[
X = \{ x \in R^p : l(x) - c = 0 \}
\]

is a \((p-q)\)-dimensional linear manifold in \( R^p \).

**Proof:** We have to show that

1. There is a \((p-q)\)-dimensional linear subspace \( U \) of \( R^p \) such that \( x, y \in X \Rightarrow x - y \in U \).

2. \( x_0 \in X, u \in U \Rightarrow x_0 + u \in X \).

Proof of (1): \( l(x) = c, l(y) = c \Rightarrow l(x - y) = 0 \). Thus \( x - y \in Ker(l) \), the kernel of the linear map \( l \). \( Ker(l) \) is a linear subspace of \( R^p \) of dimension \( p - q \).

Proof of (2): \( x_0 \in X, u \in U \Rightarrow l(x_0 + u) = l(x_0) = c \).

Let us now return to the minimum distance problem. For given \( x_1, \ldots, x_n \in R^p \), we want to find a linear map \( l : R^p \to R^k \) and a vector \( c \in R^k \), such that

\[
L = \{ x \in R^p : l(x) - c = 0 \}
\]

is the closest linear manifold of co-dimension \( k \) to \( x_1, \ldots, x_n \). We already showed that \( Ker(l) \) is the space spanned by the \( p - k \) largest eigenvectors of \( \Sigma = 1/n \sum (x_i - \bar{x})(x_i - \bar{x})^T \). Thus, if \( l(z) = Az \), the row space of \( A \) has to be spanned by the \( k \) smallest eigenvectors of \( \Sigma \). We also know that \( L \) has to pass through \( \bar{x} \Rightarrow c = l(\bar{x}) \).

**Summary:**
A k-dimensional linear manifold $L$ with associated subspace $U$ can be represented in parametric form

$$L = \{ \mathbf{x} \in \mathbb{R}^p : \mathbf{x} = \mathbf{u}_0 + \sum_{i=1}^{k} x_i \mathbf{u}_i \}$$

with $\mathbf{u}_0 \in L$ and $\mathbf{u}_1, \ldots, \mathbf{u}_k$ a basis for $U$. Alternatively, $L$ can be represented as the null space of an affine map $\mathbf{x} \to l(\mathbf{x}) - \mathbf{c}$, where $l : \mathbb{R}^p \to \mathbb{R}^{p-k}$ is a linear map of full rank and $\mathbf{c} \in \mathbb{R}^{p-k}$.

For given $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$, let $L$ denote a linear manifold closest to $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then $\mathbf{\bar{x}} \in L$, and

- $[\mathbf{u}_1, \ldots, \mathbf{u}_k]$ is spanned by $k$ largest eigenvectors of $\Sigma$.
- $l(z) = A\mathbf{z}$, with the row space of $A$ spanned by $k$ smallest eigenvectors of $\Sigma$, and $\mathbf{c} = A\mathbf{\bar{x}}$. 