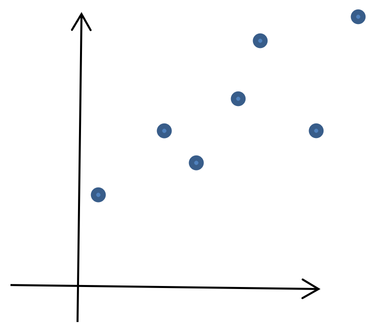
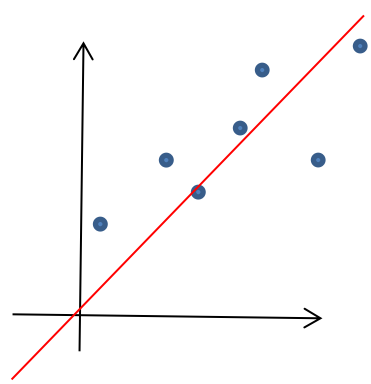
§ 6. Приложения сингулярного разложения

Мы уже обсуждали, как решать несовместные системы уравнений. Понятно, что в жизни в основном возникают именно такие.

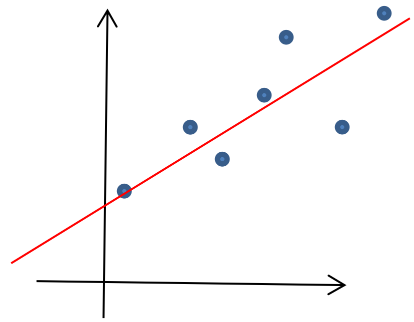
Сейчас мы поговорим немного о другом. Пусть у вас есть данные о каких-то объектах. Эти данные представлены численно тем или иным образом. Иными словами, каждый объект описывается вектором в соответствующем многомерном пространстве. Если характеристик только две, то можно считать, что множество объектов – это точки на плоскости, если три, то это точки в пространстве. Как правило, их гораздо больше. Но пока посмотрим на двухпараметрическое множество.



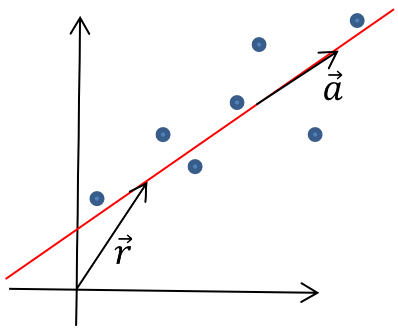
Мы пытались найти линейное подпространство (в данном случае прямую, проходящую через начало координат), чтобы точки как можно меньше от него уклонялись.



Но видно, что такая прямая не очень хороша. Лучше было бы так:



? Как задаётся такая прямая?



Она задается как множество векторов $\vec{r} + < \vec{a} > = \{ \vec{r} + \vec{x} \mid \vec{x} \in < \vec{a} > \}$.

В общем случае мы рассматриваем множество $r + M$, где r – это элемент пространства, в котором расположились наши данные, а M – подпространство, которое лучше всего приближает наши данные.

Определение 2. Множество $r + M$, где r – это элемент пространства L, а M – его подпространство, называется **линейным многообразием**. Размерность пространства M называется **размерностью** этого многообразия.

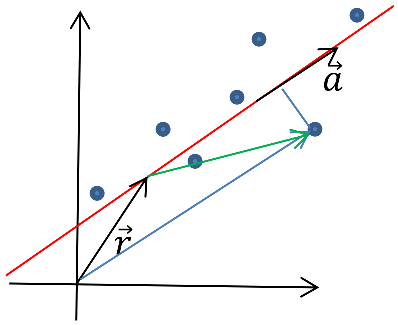
Лемма 1. $r + M = s + M$ тогда и только тогда, когда $r - s \in M$.

Доказать самостоятельно.

Прежде всего, научимся находить вектор r.

Теорема 1. Пусть для любых векторов $x_1, x_2, \ldots, x_n$ линейное многообразие $r + M$ размерности k таково, что для любого многообразия $r' + M'$ размерности не больше k выполняется неравенство $\sum_{i=1}^{n} d(x_i,\ r + M)^2 \leq \sum_{i=1}^{n} d(x_i,\ r' + M')^2$. Тогда $r + M = \dfrac{x_1 + x_2 + \cdots + x_n}{n} + M$.

Доказательство. Заметим, что $d(x_i, r' + M')$ – это длина ортогональной составляющей при проектировании элемента $x_i - r'$ на подпространство M'.



Обозначим для краткости $\dfrac{x_1+x_2+\cdots+x_n}{n}$ буквой s и рассмотрим линейное многообразие $s + M$.

Обозначим через w_i ортогональную составляющую вектора $x_i - s$.

Поскольку $\sum_{i=1}^{n}(x_i - s) = 0$, то и $\sum_{i=1}^{n} w_i = 0$. Тогда $w_n = -\sum_{i=1}^{n-1} w_i$.

Пусть t – ортогональная составляющая элемента $s - r$. Так как $x_i - r = (x_i - s) + (s - r)$, ортогональная составляющая элемента $x_i - r$ равна $w_i + t$.

Значит,

$$\sum_{i=1}^{n} d(x_i,\ r+M)^2 = \sum_{i=1}^{n} |w_i + t|^2 = \sum_{i=1}^{n-1} |w_i + t|^2 + |w_n + t|^2 =$$

$$= \sum_{i=1}^{n-1} |w_i + t|^2 + \Big|t - \sum_{i=1}^{n-1} w_i\Big|^2 =$$

$$= \sum_{i=1}^{n-1}\big(|w_i|^2 + |t|^2 + (w_i,t) + (t,w_i)\big) + |t|^2 - \sum_{i=1}^{n-1}\big((w_i,t) + (t,w_i)\big) + \Big|\sum_{i=1}^{n-1} w_i\Big|^2 =$$

$$= \sum_{i=1}^{n} |w_i|^2 + n|t|^2 \geq \sum_{i=1}^{n} |w_i|^2 = \sum_{i=1}^{n} d(x_i,\ s+M)^2 \geq \sum_{i=1}^{n} d(x_i,\ r+M)^2.$$

Следовательно, $t = 0$, а это означает, что $s - r \in M$, т.е. $r + M = s + M$. □