Suppose we have n p-dimensional data points $\{x_i\}: i=1\ldots n$. Can we find a point x_0 that is representative of the whole data?

- Minimise a squared error criterion

$$J_0(x_0) = \sum_{i=1}^{n} \|x_0 - x_i\|^2$$

- Solution given by sample mean

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

"Zero dimensional" representation - doesn't capture any variability in the data!

One dimensional representation more useful - project the data onto a line running through the mean.

Let $\underline{u}$ be a unit vector through $\underline{m}$, ②

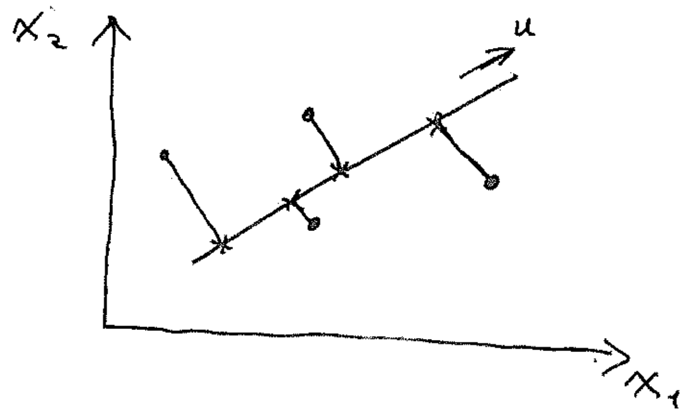
ie. $\underline{u}^T\underline{u} = 1$. (Only care about direction)

Approximate each data point x_i by

$$\hat{\underline{x}}_i = \underline{m} + a_i \underline{u}$$

where a_i is distance between x_i & the mean m.

* Least squares solution obtained by projecting x_i onto the line in direction of $\underline{u}$ that passes through the mean $\underline{m}$.



2d → 1d.

③

Need to find n coefficients $\{a_i\}$ that minimise,

$$J_1(a_1 \ldots a_n; \underline{u}) = \sum_{i=1}^{n} \|(m + a_i u) - x_i\|^2$$

$$= \sum_{i=1}^{n} \|a_i u - (x_i - m)\|^2$$

$$= \sum_{i=1}^{n} a_i^2 \|u\|^2 - 2 \sum_{i=1}^{n} a_i u^T (x_i - m)$$

$$+ \sum_{i=1}^{n} \|x_i - m\|^2 .$$

Differentiate wrt a_i & set to zero,

$$a_i = u^T (x_i - m) .$$

Sub. back in above to get

$$J_1(a_1, \ldots a_n; u) = \sum_{i=1}^{n} a_i^2 - 2 \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} \|x_i - m\|^2$$

$$= - \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} \|x_i - m\|^2 .$$

$\therefore$ we can find $\{a_i\}$s .

What about optimal direction $\underline{u}$? 4

Rewrite J in terms of $\underline{u}$.

$$J_1(u) = -\sum_{i=1}^{n} \left[u^T(x_i - m)\right]^2 + \sum_{i=1}^{n} \|x_i - m\|^2$$

$$= -\sum_{i=1}^{n} u^T(x_i - m)(x_i - m)^T u + \sum_{i=1}^{n} \|x_i - m\|^2$$

$$= -u^T S u + \underbrace{\sum_{i=1}^{n} \|x_i - m\|^2}_{\text{constant}}$$

where $S = \sum_{i=1}^{n} (x_i - m)(x_i - m)^T$.

$\Rightarrow$ Subject to $\|u\| = 1$, want to find

$$\min_u J_1(u) = \max u^T S u$$

Use Lagrange multiplier, find max of

$$u^T S u - \lambda(u^T u - 1)$$

Differentiate wrt u & set equal to zero

$$2Su - 2\lambda u = 0 \quad \therefore \quad Su = \lambda u.$$

So, u must be eigenvector of S.

Maximise $u^T S u$ by selecting eigenvector of S corresponding to largest eigenvalue since

$$u^T S u = \lambda u^T u = \lambda .$$

More generally, we could seek an optimal q-dimensional representation

$$\hat{x}_i = m + \sum_{j=1}^{q} a_{ij} u_j \qquad q < p .$$

$$J_q(u_1 \ldots u_q) = \sum_{i=1}^{n} \| (m + \sum_{j=1}^{q} a_{ij} u_j) - x_i \|^2$$

Minimised when $u_1 \ldots u_q$ are the eigenvectors of S taken in decreasing order of their eigenvalues!

PCA Analysis

1. Collect all data into $n \times p$ matrix,
   $X = (x_1 \dots x_n)^T$

2. Transform the data to have zero mean.

3. Find p eigenvector/value pairs,
   $u_1 \dots u_q$, $\lambda_1 \dots \lambda_q$ $\quad q < p$.

4. Set $Y \approx \overset{n \times p}{X} W$

   Y: $n \times q$ approximation

   W: $p \times q$ with columns corr. to the eigenvectors