Principal component analysis I: one-dimensional projection

Sanjoy Dasgupta

University of California, San Diego

Topics we'll cover

1. The variance of a projection
2. Finding the best single direction

The best direction

Suppose we need to map our data $x \in \mathbb{R}^d$ into just **one** dimension:

$$x \mapsto u \cdot x \quad \text{for some unit direction } u \in \mathbb{R}^d$$

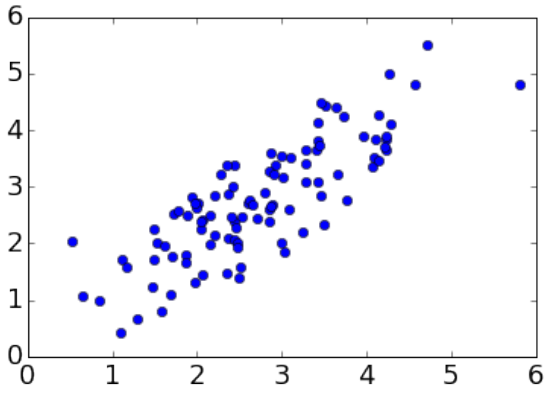
What is the direction u of maximum variance?

Useful fact 1:

- Let Σ be the $d \times d$ covariance matrix of X.
- The variance of X in direction u (the variance of $X \cdot u$) is:

$$u^T \Sigma u.$$

Best direction: example



Here covariance matrix $\Sigma = \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}$

The best direction

Suppose we need to map our data $x \in \mathbb{R}^d$ into just **one** dimension:

$$x \mapsto u \cdot x \quad \text{for some unit direction } u \in \mathbb{R}^d$$

What is the direction u of maximum variance?

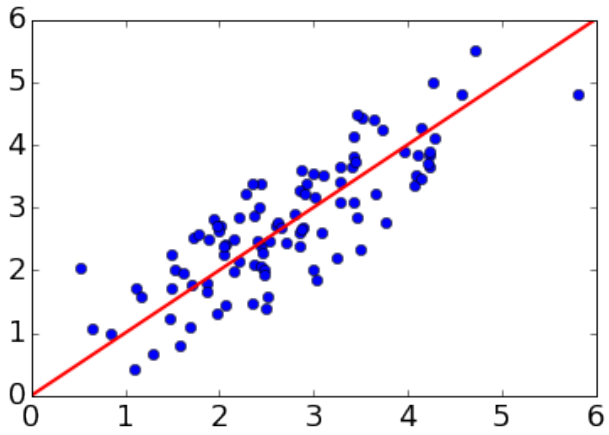
Useful fact 1:

- Let Σ be the $d \times d$ covariance matrix of X.
- The variance of X in direction u is given by $u^T \Sigma u$.

Useful fact 2:

- $u^T \Sigma u$ is maximized by setting u to the first **eigenvector** of Σ.
- The maximum value is the corresponding **eigenvalue**.

Best direction: example



Direction: **first eigenvector** of the 2×2 covariance matrix of the data.

Projection onto this direction:
the top **principal component** of the data