

Linear projections

Sanjoy Dasgupta

University of California, San Diego

Topics we'll cover

- ① Why dimensionality reduction?
- ② Choosing informative coordinates
- ③ Projecting onto an informative direction

Dimensionality reduction

Why reduce the number of features in a data set?

- ① It reduces storage and computation time.
- ② High-dimensional data often has a lot of redundancy.
- ③ Remove noisy or irrelevant features.

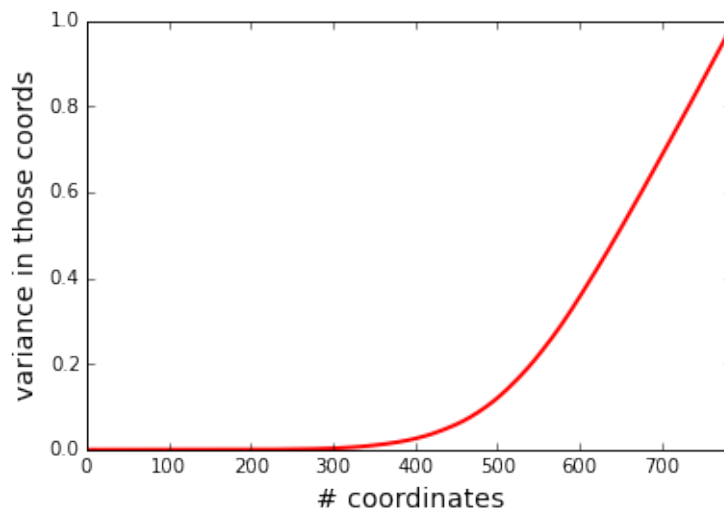
Example: are all the pixels in an image equally informative?



If we were to choose a few pixels to discard, which would be the prime candidates?

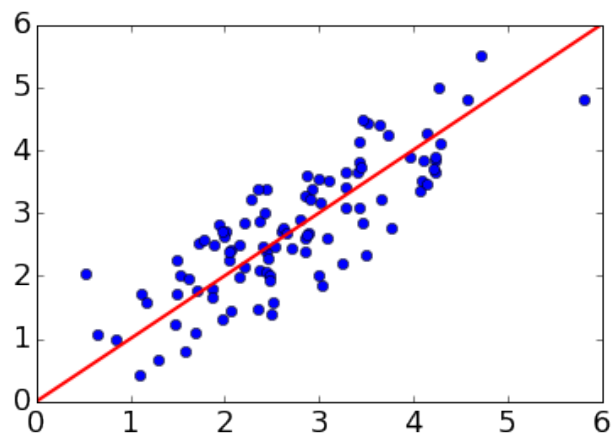
Eliminating low variance coordinates

MNIST: what fraction of the total variance lies in the 100 (or 200, or 300) coordinates with lowest variance?



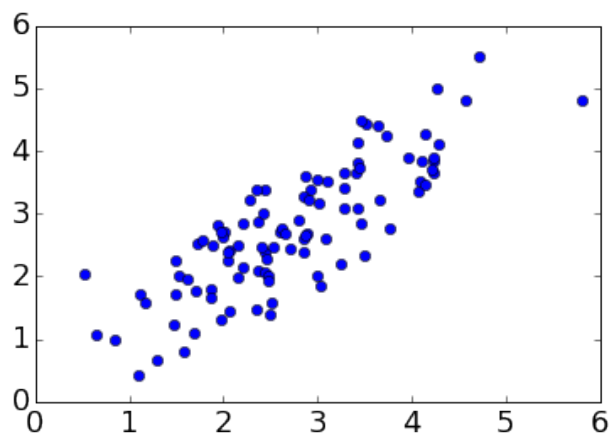
The effect of correlation

Suppose we wanted just one feature for the following data.



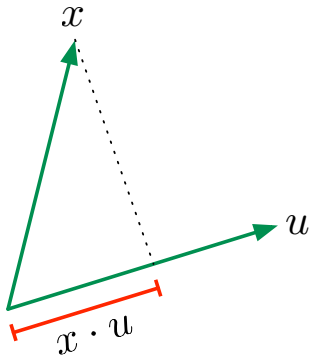
This is the **direction of maximum variance**.

Comparing projections



Projection: formally

What is the projection of $x \in \mathbb{R}^d$ in the **direction** $u \in \mathbb{R}^d$?
Assume u is a unit vector (i.e. $\|u\| = 1$).



Projection is

$$x \cdot u = u \cdot x = u^T x = \sum_{i=1}^d u_i x_i.$$

Examples

What is the projection of $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ along the following directions?

- ① The x_1 -axis?
- ② The direction of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$?