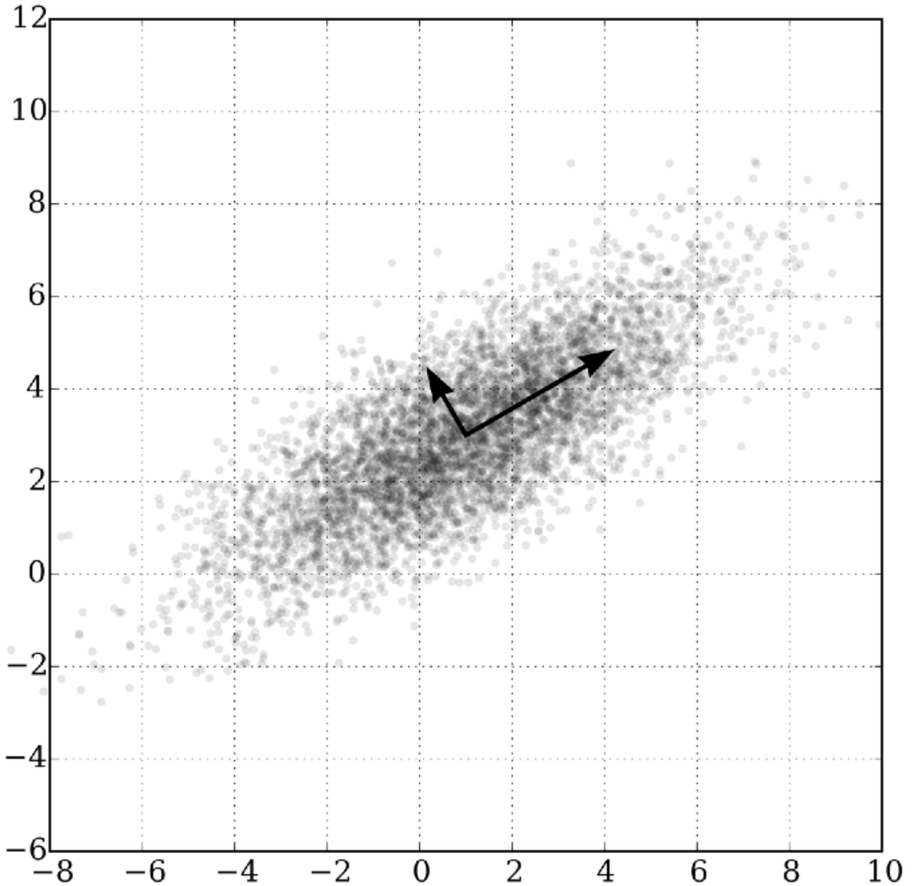
Principal component analysis (PCA)

- An *automated* way to do dimensionality reduction
- A linear transformation of coordinates in decreasing order of variance
  - First principal component has the largest variance
  - Second principal component has second largest variance
  - And so forth
- Dimensions can be reduced by keeping the highest-variance dimensions

- See https://en.wikipedia.org/wiki/Principal_component_analysis



[2]

The matrices of PCA

- **CV** = **PV**
- **C**: covariance matrix
  - C_{kl} is the covariance between dimensions k and l.
  - Usually empirically estimated from data.
- **V**: matrix of column *eigenvectors*
  - V_{kl} is the
    - importance of the original coordinate k
    - in the transformed coordinate l.
    - known as the principal components
    - the columns are orthonormal vectors
- **P**: diagonal matrix of *eigenvalues*
  - $P_{kl} = \lambda_{kl}$ if k = l
  - $P_{kl} = 0$ otherwise
  - variances in transformed coordinate system
  - scaling of the eigenvectors
- All three matrices have the same size