Day 26

Principal Components

1. Given data matrix X with N rows (samples) and k columns (features) – assume each feature has mean zero.
2. The matrix $Q = \frac{1}{N} X^\intercal X$ is symmetric and its entries are the variances/covariances.
3. If v is a vector, then Xv is called a "score" – a synthetic measure of the data.
4. The variance of the score is $v^\intercal Q v$.
5. Critical points of variance are eigenvectors of Q.
6. These critical directions are called "principal components".