CISC 271   Class 23

PCA and the Rayleigh Quotient

Text Correspondence: $\sim$

Main Concepts:

- *Scores of original data*
- *PCA as an optimization problem*
- *Rayleigh quotient for optimization*

Sample Problem, Machine Inference: What does PCA optimize?

In previous classes, we have explored how principal components analysis (PCA) can be formulated as the solution to an approximation problem. The results that we relied on was the Eckart-Young theorem: for certain matrix norms, any matrix can be written as the sum of rank-1 matrices and the truncation of this series is the "best" approximation to the matrix that is under consideration.

23.1 Zero-Mean Data and Maximization

We suppose that our data are given as a matrix $A \in \mathbb{R}^{m\times n}$, from which we compute a zero-mean matrix

$$M = A - \vec{1}\,\bar{A} = A - \tfrac{1}{m}\vec{1}\,\vec{1}^{\,T}A$$

We will suppose that the zero-mean matrix M is a "tall thin" matrix, with $m > n$. This implies that the column space is a vector space $\mathbb{U} \subset \mathbb{R}^m$.

Consider a vector $\vec{u} \in \mathbb{R}^n$. What vector $\vec{u}$ is mapped to the "largest" column vector $M\vec{u}$? To answer this question, we need to specify what we mean by "largest".

Recall vector norms, and specifically the ℓ^2 norm that is also called the Euclidean vector norm. We write this norm as $\|\cdot\|_2$ or, informally, simply as $\|\cdot\|$. We can use this norm to create an *objective* function, which is a function $f : \mathbb{R}^m \to \mathbb{R}$ that has a vector argument and that maps to a real number. The ℓ^2 norm of $M\vec{u}$ is $\|M\vec{u}\|$; let us consider using the squared norm as our objective when we ask about the "largest" vector to which $\vec{u} \in \mathbb{R}^n$ maps.

We can formulate and expand our objective function as

$$\begin{aligned} f(\vec{u}) &= \|M\vec{u}\|^2 \\ &= [M\vec{u}]^T[M\vec{u}] \\ &= \vec{u}^T M^T M \vec{u} \\ &= \vec{u}^T S \vec{u} \end{aligned} \tag{23.1}$$

The matrix S in Equation 23.1 is the scatter matrix for the zero-mean matrix M. Finding the largest vector that M maps $\vec{u}$ to is the same as asking what vector argument maximizes the quadratic form of the scatter matrix S. This is an "argument maximum" problem, which in mathematics is often abbreviated as an *argmax* problem.

Instead of asking what vector argument $\vec{u}$ maximizes our objective function $f(\cdot)$, we usually ask what *unit* vector maximizes $f(\cdot)$. This is a *constrained* argmax problem. Suppose that the answer to this problem is a unit vector $\vec{w}$. The problem can be written as

$$\vec{w} = \underset{\vec{u}\in\mathbb{R}^n : \|\vec{u}\|=1}{\operatorname{argmax}} \ \vec{u}^T S \vec{u} \tag{23.2}$$

We can use the spectral decomposition of the scatter matrix to solve Equation 23.5. The answer is that $\vec{w}$ is the unit eigenvector of the largest eigenvalue of S, so

$$\vec{w} = \vec{v}_{\text{MAX}}(S) \tag{23.3}$$

By construction, we know that S is a symmetric positive semidefinite matrix with each eigenvalue greater than or equal to zero. Strictly speaking, any vector $\vec{w}$ that is in the eigenvector space of $\lambda_{\text{MAX}}(S)$ will be a possible solution. For empirical data, $\lambda_{\text{MAX}}(S)$ is almost invariably unique.

23.2 Rayleigh Quotient of a Symmetric Matrix

The constrained optimization problem of Equation 23.5 arises often enough that there is an unconstrained formulation – or, at least, a much less constrained formulation. The formulation is named after the mathematician and physicist John William Strutt, who is better known as the Nobel laureate Lord Rayleigh.

The Rayleigh quotient is defined as the ratio of the quadratic form $\vec{u}^T S \vec{u}$ and the squared norm of $\vec{u}$, which can be written as $\vec{u}^T\vec{u}$. As $\vec{u} \to \vec{0}$, the Rayleigh quotient approaches zero, so we can write the Rayleigh quotient as

$$R(S, \vec{u}) \stackrel{\text{def}}{=} \begin{cases} \dfrac{\vec{u}^T S \vec{u}}{\vec{u}^T \vec{u}} & \text{if} \quad \vec{u} \neq \vec{0} \\ 0 & \text{if} \quad \vec{u} = \vec{0} \end{cases} \tag{23.4}$$

We can write Equation 23.5, using Equation 23.4, as the unconstrained optimization problem

$$\vec{w} = \operatorname*{argmax}_{\vec{u} \in \mathbb{R}^n} R(S, \vec{u}) \tag{23.5}$$

The solution to Equation 23.5 is the same as that of Equation 23.2. Solutions to the "argument maximum" and the "maximum" problem of the Rayleigh quotient can be succinctly stated as

$$\begin{array}{rcl} \vec{v}_{\text{MAX}} & = & \text{argmax}_{\vec{u} \in \mathbb{R}^n} R(S, \vec{u}) \\ \lambda_{\text{MAX}} & = & \max_{\vec{u} \in \mathbb{R}^n} R(S, \vec{u}) \end{array} \tag{23.6}$$

Both formulations of Equationeq:solveRquot can be solved using the spectral decomposition of the symmetric scatter matrix S.