Derivation of the First Principal Component

Suppose we have an \( n \times p \) data matrix,

\[
X = \begin{pmatrix}
  \mathbf{x}_1 \\
  \mathbf{x}_2 \\
  \vdots \\
  \mathbf{x}_n
\end{pmatrix}
\]

where \( n \) is the number of observations and \( p \) is the number of features. Why is the direction of maximal variation the eigenvector of \( X^TX \) with the largest eigenvalue? Let’s answer this question.

1. For a vector \( \mathbf{v} \) of length 1 (\( ||\mathbf{v}|| = 1 \)), explain in words or in pictures what \( \mathbf{x}_i \cdot \mathbf{v} \) represents.

2. What does \( \sum_{i=1}^{n} (\mathbf{x}_i \cdot \mathbf{v})^2 \) measure and why?

3. Explain why the optimization problem for the first principal component is:

\[
\text{maximize } \sum_{i=1}^{n} (\mathbf{x}_i \cdot \mathbf{v})^2 \\
\text{subject to } ||\mathbf{v}||^2 = 1.
\] (1)

The problem (1) is a constrained optimization problem, so we have to use Lagrange multipliers.

4. Write down the Lagrangian \( L(\mathbf{v}, \lambda) \).
5. Show that the solution to (1) is an eigenvector of $X^T X$ by: calculating the gradient $\frac{\partial L}{\partial v}$, setting it equal to 0, and rearranging the expression to show that $v$ is an eigenvector.

So far, we have shown that the direction $v$ of maximal variation is some eigenvector of $X^T X$. That’s progress! We started with infinitely many possible directions $v$ and whittled it down to $p$ directions: the $p$ eigenvectors $v_1, ..., v_p$ of $X^T X$ with eigenvalues $\lambda_1, ..., \lambda_p$. (While we’re at it, let’s assume that each eigenvector $v_k$ has been normalized so that $||v_k|| = 1$.)

Of course, we could just try all $p$ directions and see which one gives the greatest value of $\sum_{i=1}^{n} (x_i \cdot v)^2$, the objective we are trying to maximize. But it turns out that there is a systematic way to determine which of the $p$ eigenvectors gives the greatest objective value, without trying them all.

6. Calculate the value of $\sum_{i=1}^{n} (x_i \cdot v_k)^2$, where $v_k$ is an eigenvector of $X^T X$. The objective should simplify substantially, if you use the fact that $v_k$ is an eigenvector with eigenvalue $\lambda_k$ and that $||v_k||^2 = 1$.

What does this tell you about which eigenvector will maximize $\sum_{i=1}^{n} (x_i \cdot v_k)^2$?