Доказать, что оптимальные направления PCA совпадают с собственными векторами матрицы ковариаций

Дано

- X – центрированная матрица признаков размера $n \times m$, где n – количество объектов, m – количество признаков.
- Ковариационная матрица Σ данных X – симметрическая матрица $\frac{1}{n-1}X^T X$ размера $n \times n$.

Доказательство

Так как X – центрированная матрица (средние значения признаков равны нулю) $\Rightarrow$ дисперсия вычисляется как сумма квадратов. Нам нужно найти такой единичный вектор u, который максимизирует дисперсию проекции данных Xu. Несмещенная дисперсия равна:

$$var(Xu) = \frac{1}{n-1} \cdot (Xu)^T (Xu),$$

$$var(Xu) = \frac{1}{n-1} \cdot u^T X^T X u$$

$$var(Xu) = u^T \left(\frac{1}{n-1} \cdot X^T X \right) u = u^T \Sigma u$$

Получается, нам нужно максимизировать $u^T \Sigma u$, при условии $u^T u = 1$.
Для этого введем некую функцию $L(u) = u^T \Sigma u - \lambda(u^T u - 1)$, где λ – некая константа. Второе слагаемое в этой функции «штрафует» первое, если u не является единичным вектором.
Теперь нам нужно максимизировать $L(U)$, для этого найдем градиент L по u и приравняем его к нулю:

$$\nabla_u L = \nabla_u (u^T \Sigma u) - \nabla_u (\lambda(u^T u - 1)) = 0$$

Используем правила дифференцирования матричных выражений:

- $\nabla_u (u^T \Sigma u) = 2\Sigma u$ (если Σ симметрична)
- $\nabla_u (\lambda u^T u) = 2\lambda u$

Получаем:

$$2\Sigma u - 2\lambda u = 0$$

$$\Sigma u - \lambda u = 0$$

$$\Sigma u = \lambda u$$

Уравнение $\Sigma u = \lambda u$ является определением собственного вектора и собственного значения для матрицы Σ. Это означает, что вектор u, который максимизирует дисперсию $u^T \Sigma u$ при условии $u^T u = 1$, должен быть собственным вектором ковариационной матрицы Σ.
Если u – собственный вектор, удовлетворяющий $\Sigma u = \lambda u$ и $u^t u = 1$, то дисперсия равна:

$$var(Xu) = u^T \Sigma u = u^T (\lambda u) = \lambda (u^T u) = \lambda$$

Таким образом, дисперсия проекции на собственный вектор u равна соответствующему собственному значению λ. Чтобы максимизировать дисперсию, мы должны выбрать собственный вектор u_1, соответствующий наибольшему собственному значению λ_1 матрицы Σ. Этот вектор u_1 и есть первая главная компонента.

Вторая главная компонента u_2 ищется так, чтобы максимизировать $u_2^T \Sigma u_2$ при условиях:

- $u_2^T u = 1$ (единичный вектор)
- $u_2^T u_1 = 0$ (ортогональность первой компоненте)

Снова применяя тот же метод, можно показать, что u_2 должен быть собственным вектором Σ, соответствующим второму по велечине собственному значению λ_2.
Далее этот процесс продолжается. k-ая главная компонента u_k является собственным вектором Σ, соответствующим k-ому по велечине собственному значению λ_k, и ортогональным всем предыдущим компонентам $u_1, \ldots, u_{k-1}$ (собственные векторы симметрической матрицы, соответствующие разным собственным значениям, автоматически ортогональны, а если есть кратные собственные значения, то можно выбрать ортогональный базис в соответствующем собственном подпространстве). $\Rightarrow$ чтд

Вывод

Оптимальные направления PCA, которые последовательно максимизируют дисперсию проекций данным при условии ортогональности к предыдущим, являются собственными векторами ковариационной матрицы Σ, упорядоченных по убыванию соответствующих им собственных значений – наибольшему собственному значению соответствует первая главная компонента, второму по велечине – вторая и тд.