# Analýza hlavních komponent – příklad

Bylo provedeno měření objemu šedé hmoty $x\_{1}$ (v cm3) a objemu likvoru $x\_{2}$ (v cm3) u pěti dětí. Naměřené hodnoty byly zaznamenány do matice $X$:

$$X=\left[\begin{matrix}101&16\\105&18\\103&42\\98&23\\93&6\end{matrix}\right].$$

U tohoto datového souboru proveďte analýzu hlavních komponent.

**Řešení:**

U analýzy hlavních komponent potřebujeme nejprve spočítat kovarianční matici $S=\left[\begin{matrix}s\_{11}&s\_{12}\\s\_{21}&s\_{22}\end{matrix}\right]$. Pro výpočet kovarianční matice potřebujeme znát průměrný objem šedé hmoty a likvoru u $n=5$ dětí:

$\overbar{x}=\left[\begin{matrix}\frac{1}{n}\sum\_{i=1}^{n}x\_{i1}&\frac{1}{n}\sum\_{i=1}^{n}x\_{i2}\end{matrix}\right]=\left[\begin{matrix}\frac{1}{5}\left(101+105+103+98+93\right)&\frac{1}{5}\left(16+18+42+23+6\right)\end{matrix}\right]=\left[\begin{matrix}100&21\end{matrix}\right]$

Jednotlivé prvky kovarianční matice poté spočítáme následujícím způsobem:

Rozptyl objemu šedé hmoty: $s\_{11}=\frac{1}{n-1}\left(\left(x\_{11}-\overbar{x}\_{1}\right)^{2}+\left(x\_{21}-\overbar{x}\_{1}\right)^{2}+\left(x\_{31}-\overbar{x}\_{1}\right)^{2}+\left(x\_{41}-\overbar{x}\_{1}\right)^{2}+\left(x\_{51}-\overbar{x}\_{1}\right)^{2}\right)=\frac{1}{5-1}\left(\left(101-100\right)^{2}+\left(105-100\right)^{2}+\left(103-100\right)^{2}+\left(98-100\right)^{2}+\left(93-100\right)^{2}\right)=\frac{1}{4}\left(1+25+9+4+49\right)=\frac{1}{4}∙88=22$

Rozptyl objemu likvoru: $s\_{22}=\frac{1}{n-1}\left(\left(x\_{12}-\overbar{x}\_{2}\right)^{2}+\left(x\_{22}-\overbar{x}\_{2}\right)^{2}+\left(x\_{32}-\overbar{x}\_{2}\right)^{2}+\left(x\_{42}-\overbar{x}\_{2}\right)^{2}+\left(x\_{52}-\overbar{x}\_{2}\right)^{2}\right)=\frac{1}{5-1}\left(\left(16-21\right)^{2}+\left(18-21\right)^{2}+\left(42-21\right)^{2}+\left(23-21\right)^{2}+\left(6-21\right)^{2}\right)=\frac{1}{4}\left(25+9+441+4+225\right)=\frac{1}{4}∙704=176$

Kovariance objemu šedé hmoty a objemu likvoru: $s\_{12}=s\_{21}=\frac{1}{n-1}\left(\left(x\_{11}-\overbar{x}\_{1}\right)\left(x\_{12}-\overbar{x}\_{2}\right)+\left(x\_{21}-\overbar{x}\_{1}\right)\left(x\_{22}-\overbar{x}\_{2}\right)+\left(x\_{31}-\overbar{x}\_{1}\right)\left(x\_{32}-\overbar{x}\_{2}\right)+\left(x\_{41}-\overbar{x}\_{1}\right)\left(x\_{42}-\overbar{x}\_{2}\right)+\left(x\_{51}-\overbar{x}\_{1}\right)\left(x\_{52}-\overbar{x}\_{2}\right)\right)=\frac{1}{5-1}\left(\left(101-100\right)\left(16-21\right)+\left(105-100\right)\left(18-21\right)+\left(103-100\right)\left(42-21\right)+\left(98-100\right)\left(23-21\right)+\left(93-100\right)\left(6-21\right)\right)=\frac{1}{4}\left(-5-15+63-4+105\right)=\frac{1}{4}∙144=36$

Kovarianční matice je tedy: $S=\left[\begin{matrix}s\_{11}&s\_{12}\\s\_{21}&s\_{22}\end{matrix}\right]=\left[\begin{matrix}22&36\\36&176\end{matrix}\right]$.

Nyní spočítáme vlastní čísla a vlastní vektory kovarianční matice – tzn., spočítáme následující determinant: $\left|\begin{matrix}22-λ&36\\36&176-λ\end{matrix}\right|$

Vypočteme charakteristický polynom: $\left(22-λ\right)∙\left(176-λ\right)-36^{2}=3872-176λ-22λ+λ^{2}-1296=λ^{2}-198λ+2576$

A jeho kořeny, které odpovídají vlastním číslům:

$λ\_{1}=\frac{198+\sqrt{\left(-198\right)^{2}-4∙1∙2576}}{2∙1}=\frac{198+170}{2}=184$

$λ\_{2}=\frac{198-\sqrt{\left(-198\right)^{2}-4∙1∙2576}}{2∙1}=\frac{198-170}{2}=14$

Následně spočítáme vlastní vektor $v\_{1}$ odpovídající prvnímu vlastnímu číslu $λ\_{1}=184$:

 $\left[\begin{matrix}22-184&36\\36&176-184\end{matrix}\right]\~\left[\begin{matrix}-162&36\\36&-8\end{matrix}\right]\~\left[\begin{matrix}-162&36\\-162&36\end{matrix}\right]\~\left[\begin{matrix}-4,5&1\\0&0\end{matrix}\right]$

$v\_{12}=a$; $\left(-4,5\right)∙v\_{11}+v\_{12}=0$ → $v\_{11}=\frac{a}{4,5}$; např. pro $a=4,5$ pak dostáváme: $v\_{1}=\left[\begin{matrix}1&4,5\end{matrix}\right]$, který je po normalizaci roven $v\_{1}=\left[\begin{matrix}\frac{1}{\sqrt{1^{2}+4,5^{2}}}&\frac{4,5}{\sqrt{1^{2}+4,5^{2}}}\end{matrix}\right]=\left[\begin{matrix}0,2169&0,9762\end{matrix}\right]$. Kontrola, že vektor má jednotkovou délku: $\left(0,2169\right)^{2}+\left(0,9762\right)^{2}=1$.

Spočítáme vlastní vektor $v\_{2}$ odpovídající druhému vlastnímu číslu $λ\_{2}=14$:

 $\left[\begin{matrix}22-14&36\\36&176-14\end{matrix}\right]\~\left[\begin{matrix}8&36\\36&162\end{matrix}\right]\~\left[\begin{matrix}36&162\\36&162\end{matrix}\right]\~\left[\begin{matrix}1&4,5\\0&0\end{matrix}\right]$

$v\_{22}=b$; $v\_{21}+4,5∙v\_{22}=0$ → $v\_{21}=-4,5b$; např. pro $b=1$ pak dostáváme: $v\_{2}=\left[\begin{matrix}-4,5&1\end{matrix}\right]$, který je po normalizaci roven $v\_{2}=\left[\begin{matrix}\frac{-4,5}{\sqrt{\left(-4,5\right)^{2}+1^{2}}}&\frac{1}{\sqrt{\left(-4,5\right)^{2}+1^{2}}}\end{matrix}\right]=\left[\begin{matrix}-0,9762&0,2169\end{matrix}\right]$. Kontrola, že vektor má jednotkovou délku: $\left(-0,9762\right)^{2}+\left(0,2169\right)^{2}=1$.

Vlastní vektory můžeme uspořádat do matice $V=\left[\begin{matrix}v\_{1}^{T}&v\_{2}^{T}\end{matrix}\right]=\left[\begin{matrix}0,2169&-0,9762\\0,9762&0,2169\end{matrix}\right]$, přičemž pořadí vlastních vektorů odpovídá pořadí vlastních čísel seřazených od největšího k nejmenšímu.

Nyní vyjádříme hlavní komponenty odpovídající vlastním číslům seřazeným od největšího k nejmenšímu – hlavní komponenty jsou lineární kombinace původních proměnných, přičemž koeficienty jsou souřadnice příslušného vlastního vektoru:

1. hlavní komponenta: $y\_{1}=0,2169∙x\_{1}+0,9762∙x\_{2}$ (pro $λ\_{1}=184$)
2. hlavní komponenta: $y\_{2}=-0,9762∙x\_{1}+0,2169∙x\_{2}$ (pro $λ\_{2}=14$)

Výpočet procent vyčerpané variability:

1. hlavní komponenta vyčerpává: $\frac{λ\_{1}}{λ\_{1}+λ\_{2}}=\frac{184}{14+184}=0,9293$ (tzn., 92,93% variability v datech)
2. hlavní komponenta vyčerpává: $\frac{λ\_{2}}{λ\_{1}+λ\_{2}}=\frac{14}{14+184}=0,0707$ (tzn., 7,07% variability v datech)

Vyčerpanou variabilitu můžeme znázornit i pomocí sutinového grafu:



Dále spočítáme korelace hlavních komponent s původními proměnnými:

 $R\left(x\_{1},y\_{1}\right)=\frac{v\_{11}∙\sqrt{λ\_{1}}}{\sqrt{s\_{11}}}=\frac{0,2169∙\sqrt{184}}{\sqrt{22}}=0,6274$

 $R\left(x\_{2},y\_{1}\right)=\frac{v\_{12}∙\sqrt{λ\_{1}}}{\sqrt{s\_{22}}}=\frac{0,9762∙\sqrt{184}}{\sqrt{176}}=0,9981$

 $R\left(x\_{1},y\_{2}\right)=\frac{v\_{21}∙\sqrt{λ\_{2}}}{\sqrt{s\_{11}}}=\frac{-0,9762∙\sqrt{14}}{\sqrt{22}}=-0,7787$

 $R\left(x\_{2},y\_{2}\right)=\frac{v\_{22}∙\sqrt{λ\_{2}}}{\sqrt{s\_{22}}}=\frac{0,2169∙\sqrt{14}}{\sqrt{176}}=0,0612$

První hlavní komponenta je vysoce korelována s objemem likvoru a středně korelována s objemem šedé hmoty. Druhá hlavní komponenta je středně záporně korelována s objemem šedé hmoty.

Na závěr vypočítáme nové souřadnice původních bodů po transformaci pomocí obou hlavních komponent spočítaných pomocí PCA:

$Y=X∙V=\left[\begin{matrix}101&16\\105&18\\103&42\\98&23\\93&6\end{matrix}\right]\left[\begin{matrix}0,2169&-0,9762\\0,9762&0,2169\end{matrix}\right]=\left[\begin{matrix}101∙0,2169+16∙0,9762&101∙\left(-0,9762\right)+16∙0,2169\\105∙0,2169+18∙0,9762&105∙\left(-0,9762\right)+18∙0,2169\\103∙0,2169+42∙0,9762&103∙\left(-0,9762\right)+42∙0,2169\\98∙0,2169+23∙0,9762&98∙\left(-0,9762\right)+23∙0,2169\\93∙0,2169+6∙0,9762&93∙\left(-0,9762\right)+6∙0,2169\end{matrix}\right]=\left[\begin{matrix}37,5&-95,1\\40,3&-98,6\\63,3&-91,4\\43,7&-90,8\\26,0&-89,5\end{matrix}\right]$

Souřadnice subjektů můžeme přímo získat i z hlavních komponent – např. pro první subjekt:

 $y\_{1}=0,2169∙x\_{1}+0,9762∙x\_{2}=0,2169∙101+0,9762∙16=37,5$

 $y\_{2}=-0,9762∙x\_{1}+0,2169∙x\_{2}=-0,9762∙101+0,2169∙16=-95,1$

Původní data i data po transformaci pomocí PCA si znázorníme:



Pokud bychom k transformaci použili pouze první vlastní vektor, získáváme data v prostoru první hlavní komponenty:

