Česká verze stránek není od roku 2013 aktualizována. Aktuální verzi této stránky najdete v její anglické verzi zde.

# PCA (Principal Component Analysis, analýza hlavních komponent)

Lineární metoda nepřímé ordinace - lineární v tomto případě znamená, že model, na kterém je metoda založená, předpokládá lineární vztah mezi odpovědí druhu a gradientem prostředí. Není vhodná na ekologická data, protože jde o metodu implicitně založenou na euklidovské vzdálenosti, která je zatížená problémem dvojitých nul (viz nahoře). V případě analýzy heterogenních dat s mnoha nulami (jak ekologická data často vypadají) je častým výsledkem tak zvaný artefakt podkovy (*horseshoe artefact*), který způsobuje, že druhově nepodobné vzorky na koncích gradientů jsou v ordinaci umístěny blízko sebe. Hlavní použití této metody při zpracování ekologických dat tedy spočívá v ordinaci matice s proměnnými prostředí, která slouží k odhalení a popsání vzájemných korelací mezi proměnnými (které v případě ekologických dat jsou velmi časté). Na druhová data je metodu možné použít také, ale za předpokladu, že jsou dostatečně homogenní, nebo v případě heterogenních dat po aplikaci tzv. Hellingerovy standardizace (více podrobností o tomto přístupu naleznete v článku Legendre & Gallagher 2001 a v následujícím vegetačním příkladě).

PCA osy (stejně jako obecně osy všech ordinačních analýz) jsou seřazené sestupně podle zachycené variability (tzv. eigenvalue), kterou vyextrahovaly z analyzovaných dat. Jakým způsobem se máme rozhodnout, kolik PCA os máme používat při interpretaci výsledků? Existuje celá řada možností, nejsnadnější jsou následující dvě (viz Borcard et al. 2011):

- Kaiser-Guttman kritérium – vypočtěte průměr všech eigenvalues a interpretujte pouze ty osy, jejichž eigenvalue je vyšší než tento průměr;
- model zlomené hole (*broken stick model*) – hůl jednotkové délky hoďte na zem tak, aby se náhodně rozpadla na počet částí rovný počtu PCA os v analýze (který odpovídá počtu druhů/proměnných) a tyto kousky seřaďte od nejdelšího po nejkratší. Proces zopakujte mnohokrát a zprůměrujte výsledky délek pro jednotlivé kousky mezi všemi pokusy (problém má samozřejmě analytické řešení, takže házet holí netřeba). Takto vzniklá distribuce délek představuje velikosti eigenvalues, které by vznikly náhodně, jenom díky tomu, že osy jsou seřazeny podle klesající variability. Jako důležité interpretujte jen ty osy, které mají eigenvalue vyšší než hodnota vygenerovaná modelem zlomené hole pro danou osu.

Obě metody mohou být vizualizovány pomocí funkce `evplot` (autorem je F. Gillet (Borcard et al. 2011), její definici najdete zde a příklad na její použití bude následovat). Model zlomené hole můžete spočítat pomocí funkce `PCAsignificance` z knihovny `BiodiversityR`.

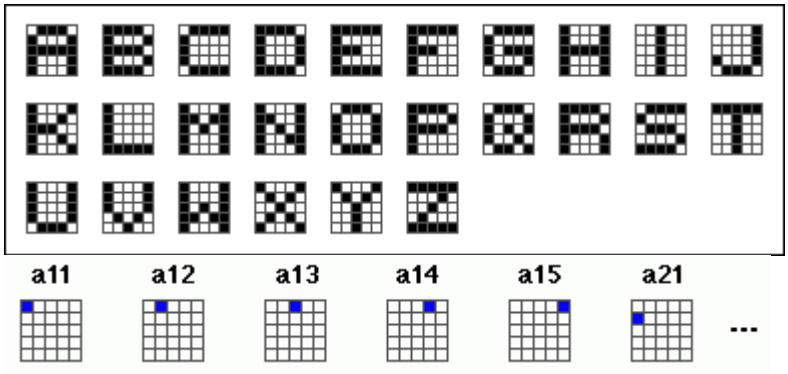
Fix Me!

---

# Cvičení

## Cvičení 1

Tento příklad není zaměřen na analýzu ekologických dat, abyste na úvod nebyli zatíženi nutností tato data intepretovat, i když je neznáte (stejně vás to ale čeká). Jde o příklad na analýzu obrazu, inspirovaný webovou stránkou François Labelle (ze které také pochází datový soubor). Podstatou je zjistit, která písmena, pokud jsou zobrazena v digitální podobě (jako mozaika černých a bílých bodů v síti 5×5 pixelů), je od sebe snadné rozeznat a které jsou snadno zaměnitelná. Zobrazení písmen a označení pixelů vypadá takto[1]:




Data jsou ke stažení v txt formátu (odděleném tabulátory) zde: letters.txt.

Nejdříve data naimportuje do eRka a pak na nich spočtěte PCA. Zajímat nás bude podobnost mezi písmeny popsaná proměnnými a11 až a55 (těchto proměnných je 25 a odkazují na černý (1) nebo bílý (0) pixel v dané pozici sítě 5×5). Nakreslete ordinační diagram (škálování 1 zaměřené na vzorky). Která písmena se od sebe liší tak, že se prakticky nedají zaměnit? Kolik PCA os potřebujeme, abychom rozdíly mezi písmeny dobře popsali (použijte broken stick model)?

Řešení cvičení 1

[1)]
http://www.cs.mcgill.ca/~sqrt/dimr/dimreduction.html

From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
**https://anadat-r.davidzeleny.net/doku.php/cs:pca**

Last update: **2017/10/11 20:36**