CCA with first three variables only

Note that exact output changes each time you run it.

Adjust the option to 0.2 and

Partitioning of mean squared contingency coefficient:

CCA(formula = community ~ spatial + env7, data = envdata)

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

---

+ env12  1 23.771 0.4569     99  0.810
+ env8   1 23.053 1.0954     99  0.390

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

---

Step: community ~ spatial

+ env9     1 47.478  1.1019     99  0.370

Start: community ~ 1

Note that exact output changes each time you run it.

Adjust the option to 0.2 and

Stepwise model fitting

---

+ env12  1 19.275 0.5525     99  0.710
+ env6   1 18.117 1.5115    999  0.219

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

---

- Step: community ~ spatial + env7 + env5

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

---

+ env3   1 19.749 0.2379     99   0.91
+ env1   1 19.672 0.3029     99   0.82

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

---

- Step: community ~ spatial + env7

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

---

vif.cca(model_fit)

4/4/2018
Tests of significance

Anova

```
Df  Chisq  F  N.Perm  Pr(>F)
Model     4  1.2771 15.2511 199 0.005 **
Residual 24  0.5024
```

Permutation test for cca under reduced model
Model: cca(formula = community ~ spatial + env7 + env5, data = envdata)

```
Df  Chisq  F  N.Perm  Pr(>F)
spatial   2  1.1099 26.5073 99 0.01 **
env7      1  0.1301  6.2121 99 0.01 **
env5      1  0.0372  1.7776 99 0.21
Residual 24  0.5024
```

```
Df  Chisq  F  N.Perm  Pr(>F)
CCA1      1  0.8242 39.3697 199 0.005 **
CCA2      1  0.4139 19.7723 199 0.005 **
CCA3      1  0.0327  1.5635 99 0.160
CCA4      1  0.0063  0.2988 99 0.860
Residual 24  0.5024
```

What to Present

Redundancy Analysis (RDA)

- CCA – aligned with CA, species relationships assumed to be unimodal
  - A CCA with no constraints = CA

- RDA – aligned with PCA, species relationships assumed to be linear
  - An RDA with no constraints = PCA

- CCA is more common for the same reasons PCA not usually used for species data.

Redundancy Analysis (RDA)

- Function rda (vegan package) works similarly to cca.

- Total inertia differs because PCA begins with variance/covariance matrix, not X

```
Inertia Proportion
Total          1.0000
Constrained    0.5533
Unconstrained  0.4467
```

```
Eigenvalues, and their contributions to the correlations

| Component | Correlation | Proportion of
correlations |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RDA1</td>
<td>0.3042</td>
<td>0.3042</td>
</tr>
<tr>
<td>RDA2</td>
<td>0.2942</td>
<td>0.2942</td>
</tr>
<tr>
<td>RDA3</td>
<td>0.2842</td>
<td>0.2842</td>
</tr>
</tbody>
</table>
```
Redundancy Analysis (RDA)

- PDA without environmental matrix = PCA

Redundancy Analysis (RDA)

- Example RDA application
  - How do various environmental variables affect morphology or behavior (plasticity question)?
  - Assumption is that you know and have measured the environmental variables that are going to impact the morphology/behavior measured.

Constrained Analysis of Principal Coordinates (capscale)

- Function `capscale` (vegan package)
- Similar to RDA but based on non-Euclidean distances
- Function works with raw data (indicate which distance metric to use) or with a distance matrix (e.g. genetic distance)
- Same `anova` function tests for significance
- Various options for dealing with data standardization
- Also called “distance based redundancy analysis” in literature

Call:
capscale(formula = community ~ spatial + env7, data = envdata, distance = "bray")

Partitioning of squared Bray distance:

<table>
<thead>
<tr>
<th>Inertia Proportion</th>
<th>Total</th>
<th>9.762</th>
<th>1.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained</td>
<td>5.993</td>
<td>0.6140</td>
<td></td>
</tr>
<tr>
<td>Unconstrained</td>
<td>3.177</td>
<td>0.3254</td>
<td></td>
</tr>
</tbody>
</table>

Important gradients are known and measured. Goal is to describe variability in species data as it relates to the measured gradients. Variability due to other unmeasured factors is ignored. Direct gradient analysis: CCA or RDA.

Gradients unknown, want to describe data or imply gradients from species data.

Data are thought to be distinctly structured. Goal is to put in distinct groups.

Cluster = hierarchical (agglomerative or divisive) or non hierarchical (k means).

Goal is to look at the relationship among samples. Indirect gradient analysis: NMDS or CA preferred. PCA or PCoA for non-community datasets.

Partial CCA

- Used when you want to control (or partial out) a variable.
  - Eg. larger spatial variable or something traditionally considered a covariate. Typically not something you’re interested in quantifying as much as controlling for.

- Recall that CCA (and RDA, capscale) use multiple regression to optimize constraints to species or sample scores.
  - Partial Regression
    - Regression coefficient of X after controlling for influence of Z on both X and Y.

---

Unconstrained NMDS approach (exploratory)

---

Wanted relationship between X (age) and y (degree of dementia). Numerous other variables confound the relationship, such as IQ and mean number of hours spent reading daily.

Residualized dementia score, effects of IQ have been "partialed" out.
Partial CCA

- All the same precautions apply as with CCA.
- Output will be nearly identical with the inclusion of the variance accounted for by the partialed variable.
- Total inertial in the ordination – same as CA
- Inertia partitioned out – comparison of the model with and without partial variable
- Inertia constrained by variables in model
- Unconstrained inertia

Partial CCA

- Uses the same function as CCA or RDA
- Code
  ```r
  cca_example <- cca(object ~ env1 + env2 + Condition(spatial), data = envdata)
  ```
- There is no significance test for the partialed variable(s), just the % variance accounted for.

Partial CCA

Test for influence of env7 while controlling for spatial variable.