Ch. 10 Regression-Based Approaches

Regression-based statistical: (1) sum of squares, (2) maximum likelihood

Regression relates a

1. Response variable: (1) presence-absence, (2) abundance, (3) biomass
2. Pre-selected set of environmental predictors

Untransformed environmental variables

Orthogonal components derived from environmental variables through multivariate analysis

Linear regression: linear model

Classical ordinary least-square (OLS) linear regression

1. Response variable is normally distributed (errors?, residuals?) - Gaussian
2. Variance does change as a function of the mean – homoscedasticity
   1. Homoscedasticity
      1. Error term
         1. Random effect in the relationship between predictors and the response variable is constant across all values of the predictor variables

Generalized linear models (GLMs)

Flexible family of regression models

Response variable to follow other distributions and non-constant variance functions

Combination of predictors (linear predictors) is related to the mean of the response variable through a link function

Transform the response variable linearly and maintain the predicted values within the original range of values allowed for the response variable

GLM

Gaussian (biomass), Poisson (species abundance, species richness), binomial (presence-absence), gamma distributions

Link-function: (1) Identity, (2) Logarithm, (3) Logit, (4) Inverse

Higher-order polynomial (transformed) regression: Response shape is not a linear function of predictors

Unimodal symmetric response: bell-curve between species abundance and environmental covariate

Non-parametric smoothing functions of predictors: GAMs, semi-parametric approaches

Applies smoothers independently to each predictor, then additively calculates the component response

Generalized linear model (GLM)

10.3 Generalized Additive Models (GAMs)

GAMs do not require postulating a shape for the response curve from a specific parametric function

Use algorithms called smoothers that automatically fit response curves – as closely as possible – to the data given the permitted level of smoothing.

GAMs – used when the relationship between the variables is expected to one of a more complex form – not easily fitted with standard parametric functions of the predictors

1. GLM with a linear or quadratic response
2. No *a priori* reason for using a particular shape

Parametric scheme:

GAM + GLM = explore general shape of the response function and then to implement it in the best possible way

R packages for GAM: (1) gam, (2) mgcv, (3) gamair, (4) GAMBoost

Package gam iteratively fits weighted additive models using backfitting – iteratively smoothing partial residuals –

1. Cubic-spline smoother
2. Collection of polynomials of degree less than or equal to 3 – defined on subintervals

Separate polynomial model is fitted to each neighborhood – moving window algorithm – fitted curve to connect all points

Predetermine the degree of smoothing applied when fitting the curve – select through cross-validation

SDMs – used degrees lower than 4 – polynomial of degree 3

Higher degrees will generate more locally complex curves

Syntax – user needs to specific a smoother – cubic spline called *s*

Degree of smoothing can change across the variables in a model – different smoothing can be specified for each variable

10.4 Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) do not require any assumptions to be made about the underlying functional relationship between species and environmental variables.

MARS fit piecewise functions that together can accommodate nonlinear responses.

Similar to GAM and smooth functions

Knots: breaks between segments and different regression lines with different slopes and thus fitted between each pair of knots

Full fitted function is constrained to have no breaks or abrupt steps.

Generalized cross-validation is used to assess the effect of adding or removing knots

Backward and forward variable selection

Mda

Earth