

SEMINARIO MODELIZACIÓN DE NICHOS ECOLÓGICOS

María Ángeles Pérez
m.angeles582@gmail.com



BRT & MaxEnt

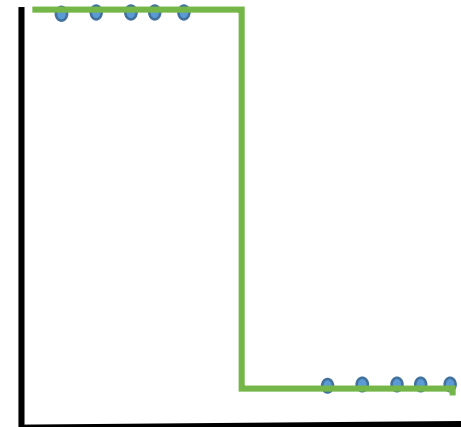
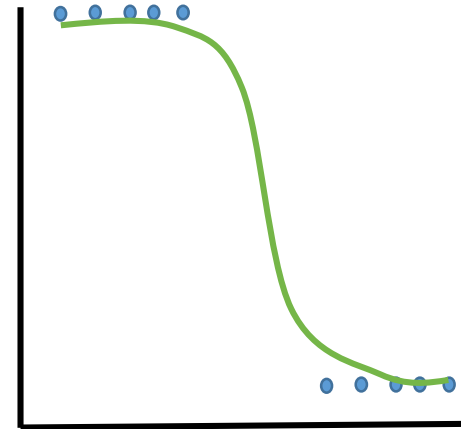
Boosted Regression Trees

Random Forest

MaxEnt

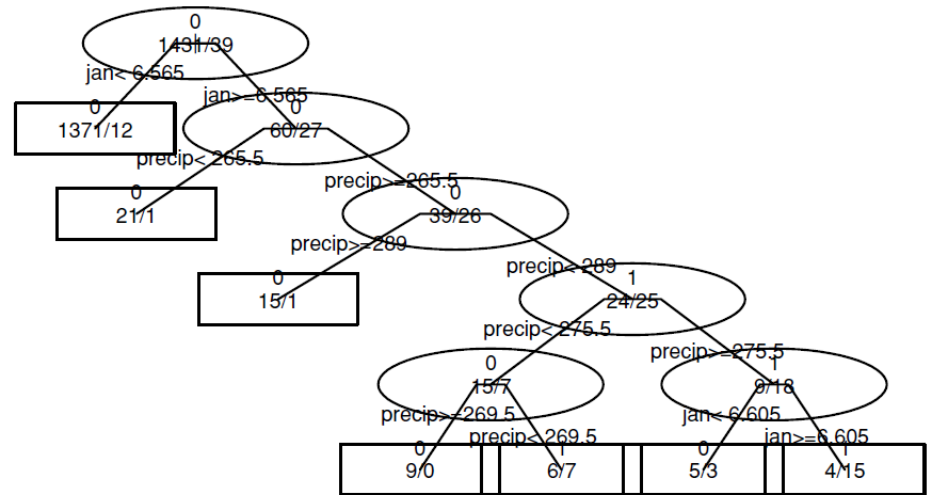
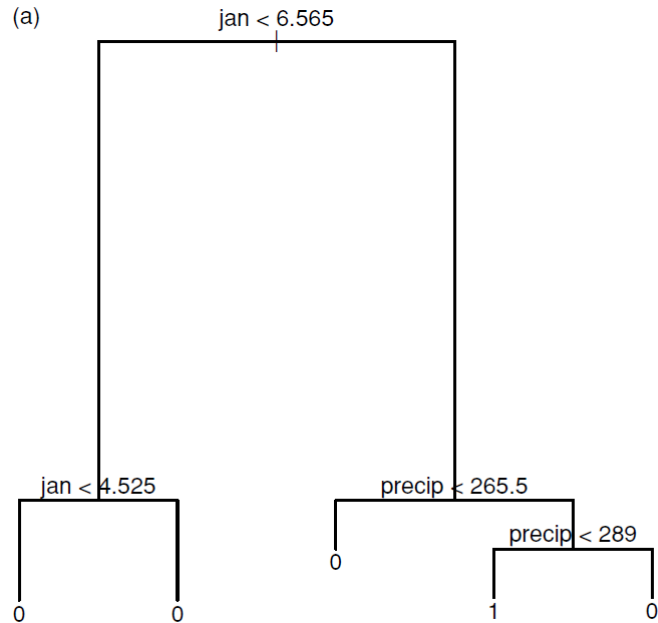
Regression Trees

- Desaparecen las curvas de suavizado
- Umbrales únicos de separación de datos
- Todas las variables son escrutadas en busca del umbral que mejor separa presencias y ausencias
- Repetición iterativa de este proceso
- ¿Cuándo paramos de dividir los datos?



Regression Trees

Descenso en la impureza del grupo



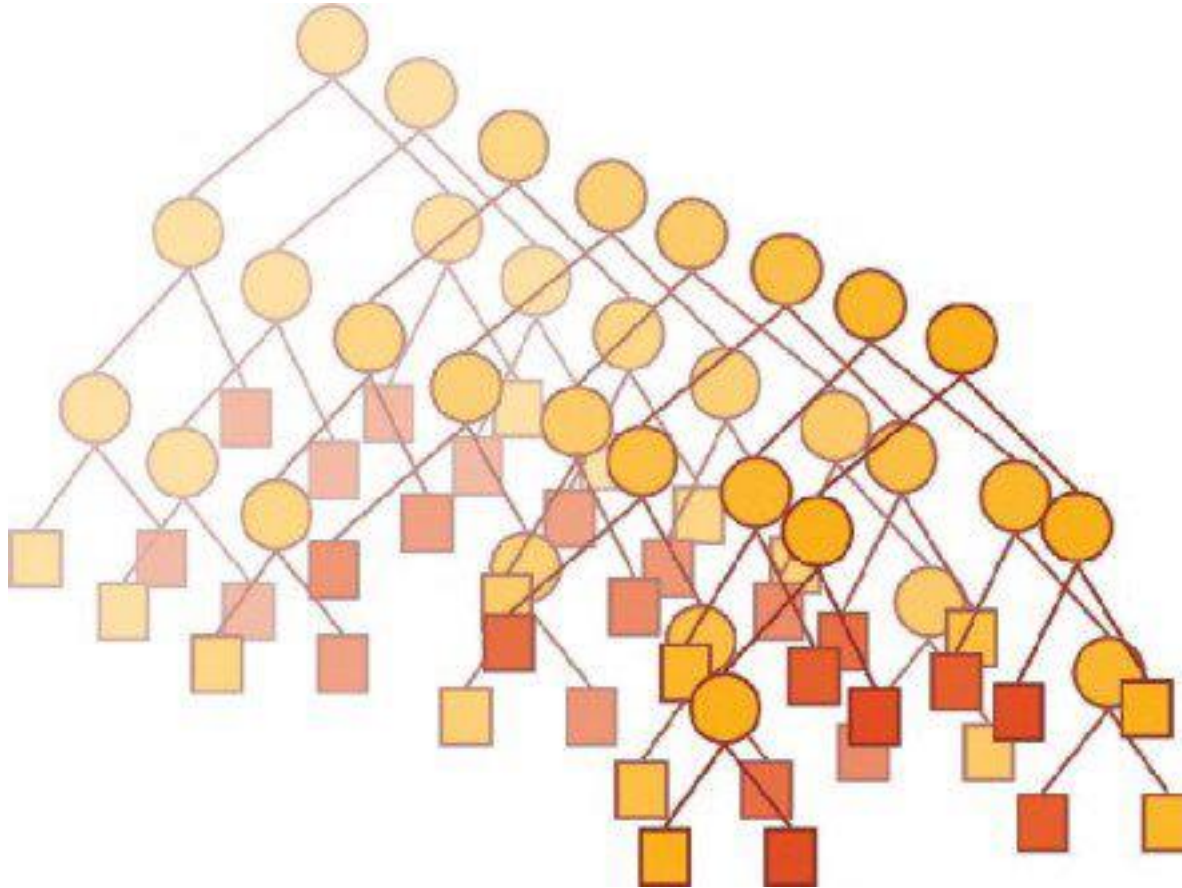
9 ausencias	6 ausencias	5 ausencias	4 ausencias
0 presencias	1 presencia	3 presencias	15 presencias

Regression Trees

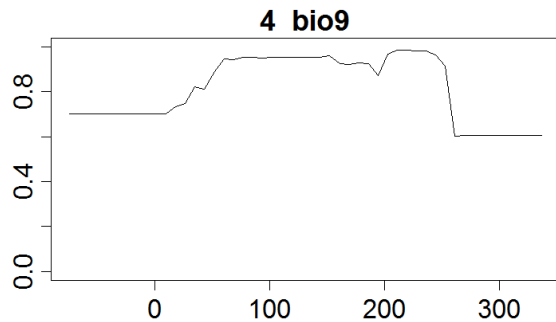
1 solo árbol es muy sensible a la eliminación o incorporación de datos



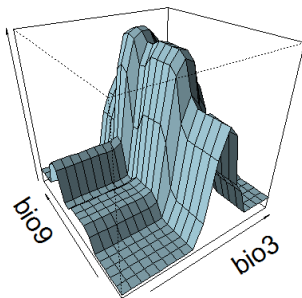
Repetimos n veces con una fracción dada de los datos!



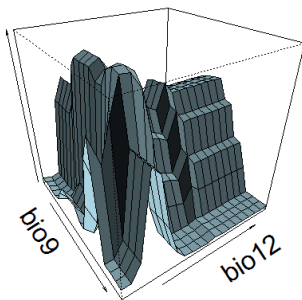
Regression Trees



2 bio3: bio9



6 bio9: bio12



Ventajas:

- Herramienta muy potente
- Buen ajuste con variables categóricas
- Analiza múltiples interacciones por regiones de rango de las variables
- Buena clasificación de importancia de variables

Desventajas:

- Necesita más registros de presencia
- Curvas difíciles de interpretar
- Potencial sobreajuste
- Caja negra

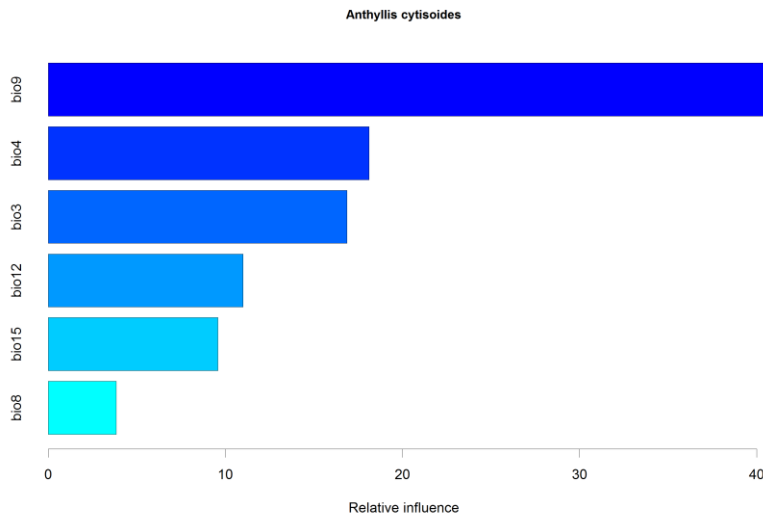
Regression Trees

Ventajas:

- Herramienta muy potente
- Buen ajuste con variables categóricas
- Analiza múltiples interacciones por regiones de rango de las variables
- Buena clasificación de importancia de variables

Desventajas:

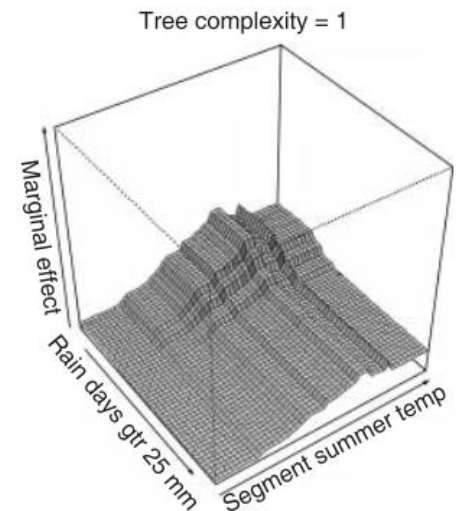
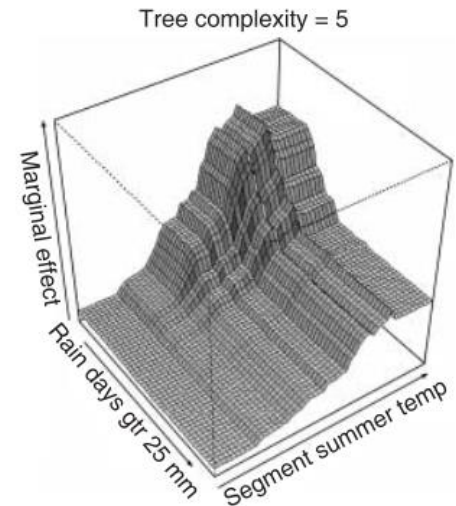
- Necesita más registros de presencia
- Curvas difíciles de interpretar
- Potencial sobreajuste
- Caja negra



Regression Trees

Boosted Regression Trees

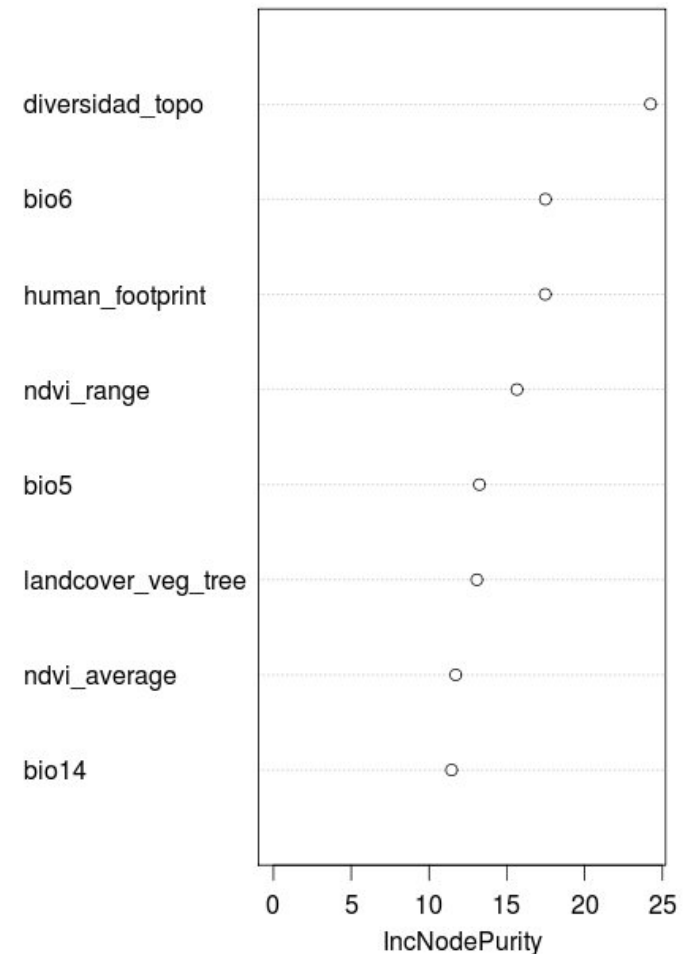
- **Lr (learning rate)** Parámetro de contracción. Determina la contribución de cada árbol al modelo creciente. Inversamente proporcional a nt
- **Tc (tree complexity)** controla las interacciones que son ejecutadas. $Tc = 1$ equivale a dos nodos terminales
- **Nt (number of trees)** Se optimiza a partir de los dos parámetros anteriores. No se recomiendan menos de 1000-2000
- Calibra cada árbol con 70% datos
- Prioriza datos difíciles de clasificar
- Ver: *A working guide to boosted regression trees*



Regression Trees

Random Forest

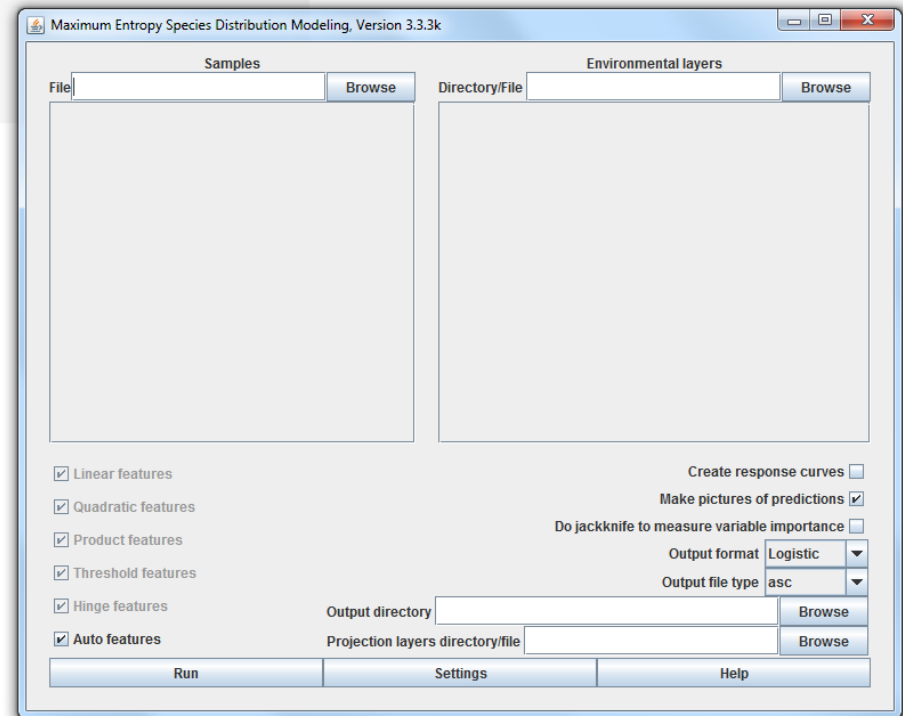
- **Library** randomForest, function randomForest
 - **ntree**: numero de arboles a calibrar
 - **mtry**: numero de variables usadas en cada arbol
 - **nodesize**: numero minimo de casos en cada nodo terminal (100!)
 - **maxnode**: numero maximo de nodos
- Terminales
- Calibra cada árbol con el 60% datos



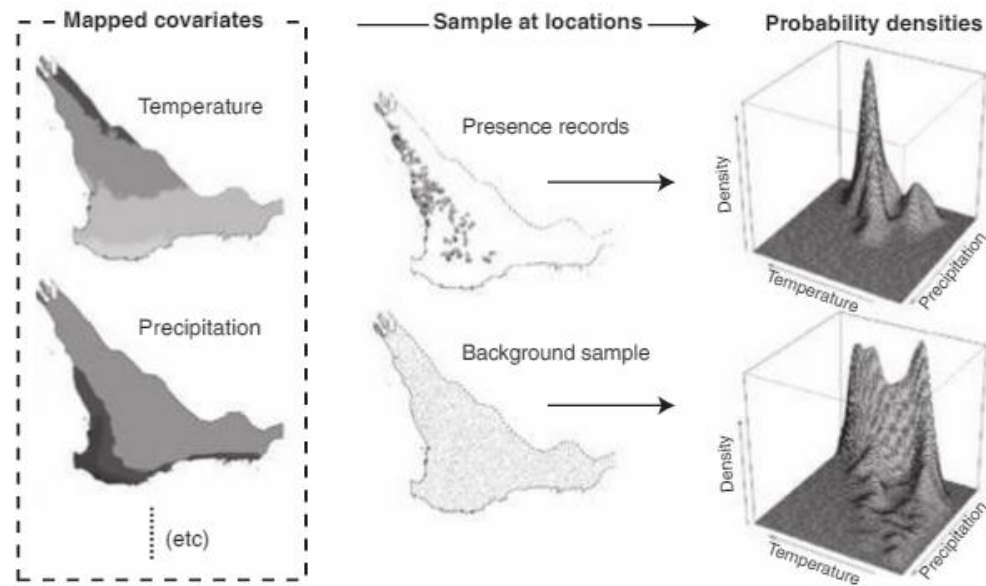
Maximum Entropy

MaxEnt

- Sencillo de usar, interfaz asequible.
- Permite trabajar con un número bajo de presencias
- Implementa eliminación de duplicados, registros sin coordenadas y puntos fuera de la extensión de trabajo
- Genera automáticamente puntos de background
- Opciones similares a los ajustes de otros algoritmos
- *Merow et al. 2013 Ecography 36: 1058–1069*



Maximum Entropy



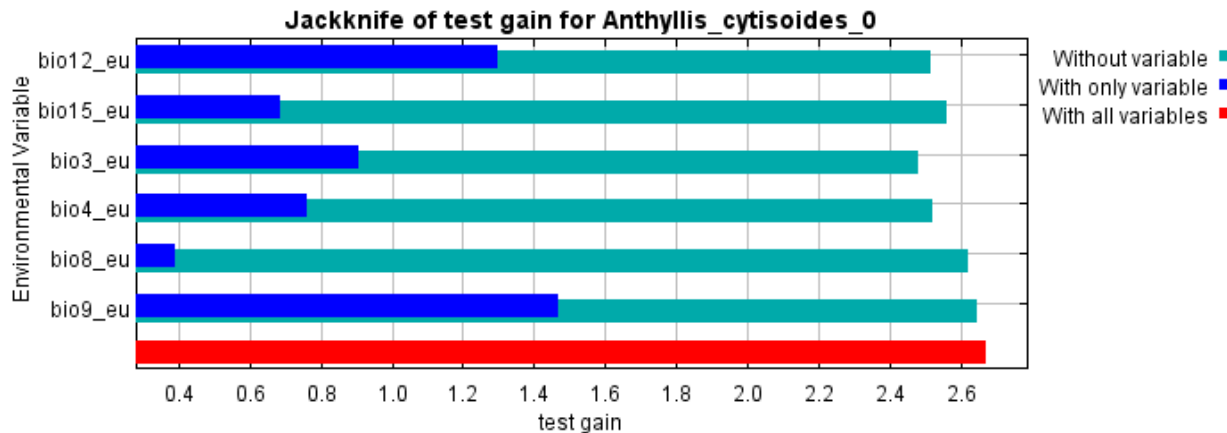
Elith et al. 2010 Diversity and Distributions, 17, 43–57

The null model for the raw distribution was the uniform distribution over the landscape, since without any data we would have no reason to think the species would prefer any location to any other

La distribución de las especies es estimada minimizando la entropía (“distancia”) entre la densidad de las variables entre los puntos de presencia y la densidad de las variables en los puntos background

Maximum Entropy

- **Background:** muestreo aleatorio de celdas. Por defecto 10000
- **Features:** variables, transformaciones y combinaciones de las variables:
- **Regularización:** valor de penalización para reducir sobreajuste. Explorar diferentes valores para optimizar el ajuste.
- **Jackknife:** Clasificación por importancia de variables. Comprueba la adecuación predictiva del modelo con y sin incluir las variables. (La importancia de las variables no tiene por qué ser constante a lo largo del territorio)

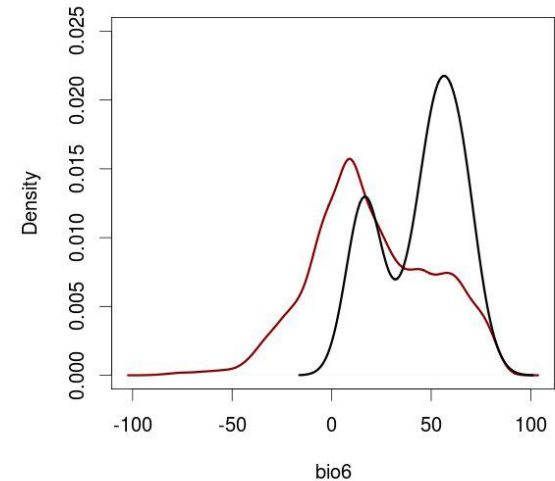


Maximum Entropy

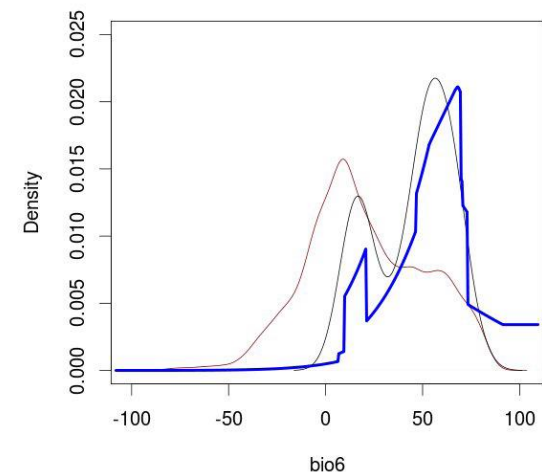
Features

- Linear: obliga a que el valor medio de una variable en la predicción coincida con su valor medio en las obsevaciones
- Quadratic: como el anterior, con la varianza
- Product: equivalente a interacción en regresión
- Threshold: transforma predictor continuo en binario mediante una step function. Ajuste similar a Decision Trees.
- Hinge: Como el anterior, pero usa funciones de suavizado. Ajuste similar a GAM

Curvas densidad
presence/background



Auto-features

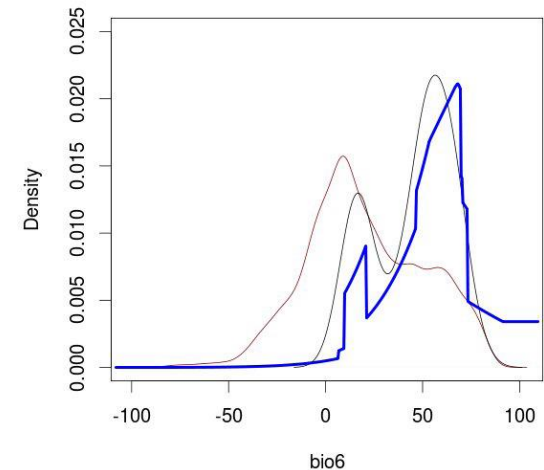


Maximum Entropy

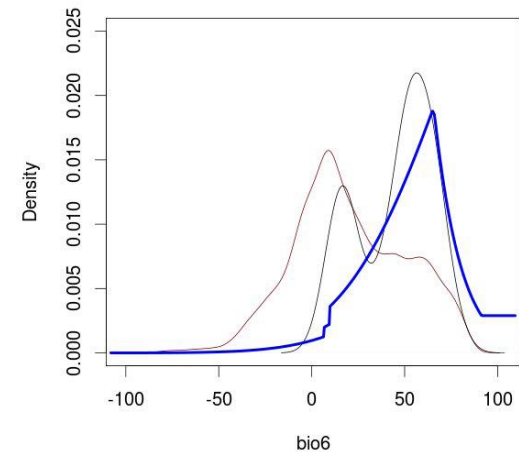
Features

- Linear: obliga a que el valor medio de una variable en la predicción coincida con su valor medio en las obsevaciones
- Quadratic: como el anterior, con la varianza
- Product: equivalente a interacción en regresión
- Threshold: transforma predictor continuo en binario mediante una step function. Ajuste similar a Decision Trees.
- Hinge: Como el anterior, pero usa funciones de suavizado. Ajuste similar a GAM

Auto-features



Auto-features + regularization =3

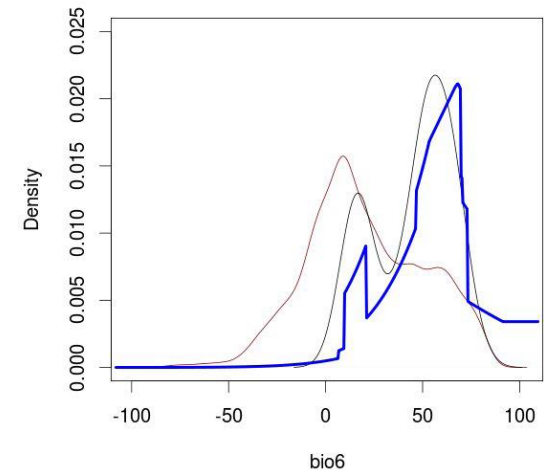


Maximum Entropy

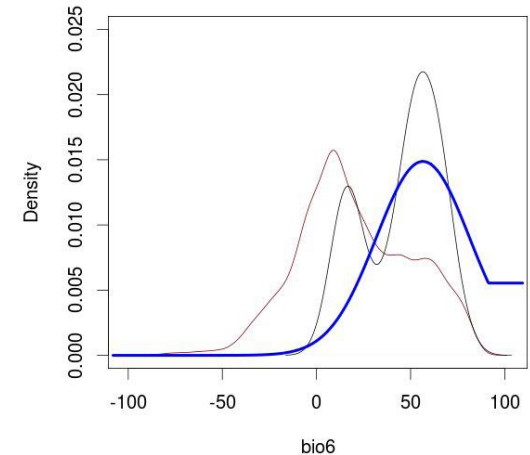
Features

- Linear: obliga a que el valor medio de una variable en la predicción coincida con su valor medio en las obsevaciones
- Quadratic: como el anterior, con la varianza
- Product: equivalente a interacción en regresión
- Threshold: transforma predictor continuo en binario mediante una step function. Ajuste similar a Decision Trees.
- Hinge: Como el anterior, pero usa funciones de suavizado. Ajuste similar a GAM

Auto-features



No hinge, no threshold

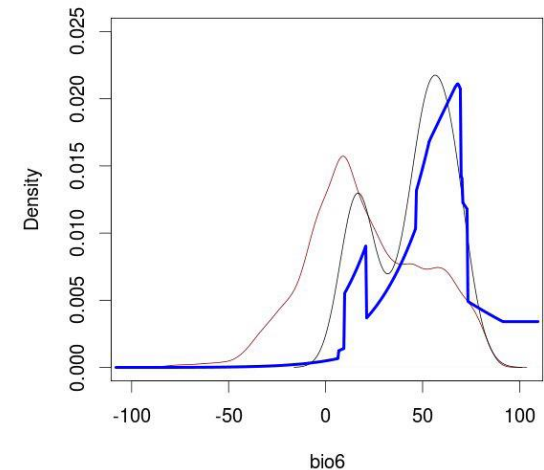


Maximum Entropy

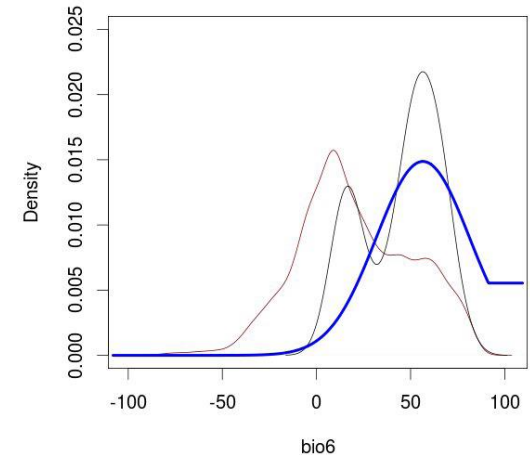
Features

- Linear: obliga a que el valor medio de una variable en la predicción coincida con su valor medio en las obsevaciones
- Quadratic: como el anterior, con la varianza
- Product: equivalente a interacción en regresión
- Threshold: transforma predictor continuo en binario mediante una step function. Ajuste similar a Decision Trees.
- Hinge: Como el anterior, pero usa funciones de suavizado. Ajuste similar a GAM
- Logistic output transforma raw output asumiendo prevalencia de 50%
- Curvas de respuesta con y sin interacción de otras variables

Auto-features



No hinge, no threshold



This is all!

GRACIAS

