

{desafío}
latam_

Modelos Aditivos Generalizados _



Motivación

Problemas recurrentes en modelación

- Nuestro modelo puede concurrir en sobreajuste o subajuste.
- Tenemos herramientas a nuestra disposición para resolver sobreajuste.
- Hasta el momento, no tenemos herramientas a nuestra disposición para resolver subajuste.
- Una forma específica de subajuste es el hecho que nuestras funciones proyectadas pueden que sean no lineales.

Soluciones a la no linealidad

Lineal

La forma funcional presenta alto sesgo

Polinomial

Reduce la capacidad de generalización del modelo

GLM

Flexibiliza el proceso de generación de datos, pero sigue siendo lineal

Solución propuesta por GAM

$$g(\mathbb{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

- La forma general de GAM busca apoyarse en las fortalezas y debilidades de cada uno de los modelos anteriores.

Solución propuesta por GAM

$$g(\mathbb{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

- La forma general de GAM busca apoyarse en las fortalezas y debilidades de cada uno de los modelos anteriores.
- Mediante la inclusión la función de vínculo, podemos flexibilizar el comportamiento de nuestro vector objetivo.

Solución propuesta por GAM

$$g(\mathbb{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

- La forma general de GAM busca apoyarse en las fortalezas y debilidades de cada uno de los modelos anteriores.
- Mediante la inclusión la función de vínculo, podemos flexibilizar el comportamiento de nuestro vector objetivo.
- La combinación lineal de parámetros nos asegura cierta estabilidad en la identificación de los resultados.

Solución propuesta por GAM

$$g(\mathbb{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

- La forma general de GAM busca apoyarse en las fortalezas y debilidades de cada uno de los modelos anteriores.
- Mediante la inclusión la función de vínculo, podemos flexibilizar el comportamiento de nuestro vector objetivo.
- La combinación lineal de parámetros nos asegura cierta estabilidad en la identificación de los resultados.
- La función de identidad específica permite flexibilizar el comportamiento de splines en cada parámetro.

Implementación y entrenamiento

Entrenamiento de GAM

$$g(\mathbb{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

Objetivo de GAM: Estimar las funciones de identidad.

Hay dos elementos a considerar en la obtención de las funciones de identidad:

- ¿Cómo se obtienen?
- ¿Cómo nos aseguramos que sean óptimas?

Obtención de la función de identidad

$$g(\mathbb{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$

La obtención de la función de identidad se realiza mediante backfitting.

Backfitting:

- Asumamos que tenemos un parámetro no identificable (no existe solución única).
- Igualamos el parámetro al promedio de observaciones en y.
- Probamos una función de suavización (Gaussian Kernel, etc...) en x para estimar una actualización de la identidad.
- Actualizamos el parámetro en base a este punto.
- Iteramos hasta que el comportamiento de f se estabilice, o se alcance algún criterio de tolerancia.

Optimización de GAM

$$\operatorname{argmin}_{\hat{f}_i} \sum_{i=1}^n (y_i - \sum_{j=1}^p f_j(x_{i,j}))^2 + \lambda \sum_j^p \int f_j''(t)^2 dt$$

1. Evaluamos el puntaje predicho a nivel de cada función de identidad generada para una observación.

Optimización de GAM

$$\operatorname{argmin}_{\hat{f}_i} \sum_{i=1}^n (y_i - \sum_{j=1}^p f_j(x_{i,j}))^2 + \lambda \sum_j^p \int f_j''(t)^2 dt$$

1. Evaluamos el puntaje predicho a nivel de cada función de identidad generada para una observación
2. Evaluamos la desviación entre lo predicho y lo observado a nivel de muestra de entrenamiento.

Optimización de GAM

$$\operatorname{argmin}_{\hat{f}_i} \sum_{i=1}^n (y_i - \sum_{j=1}^p f_j(x_{i,j}))^2 + \lambda \sum_{j=1}^p \int f_j''(t)^2 dt$$

1. Evaluamos el puntaje predicho a nivel de cada función de identidad generada para una observación
2. Evaluamos la desviación entre lo predicho y lo observado a nivel de muestra de entrenamiento.
3. Regularizamos la suavización de la función (trade off entre una penalización y un spline propuesto por el algoritmo).

Optimización de GAM

$$\underset{\hat{f}_i}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \sum_{j=1}^p f_j(x_{i,j}))^2 + \lambda \sum_j^p \int f_j''(t)^2 dt$$

1. Evaluamos el puntaje predicho a nivel de cada función de identidad generada para una observación
2. Evaluamos la desviación entre lo predicho y lo observado a nivel de muestra de entrenamiento.
3. Regularizamos la suavización de la función (trade off entre una penalización y un spline propuesto por el algoritmo).
4. Evaluamos aquella función candidata que minimice este error cuadrático regularizado.

Interpretación

Dependencia Parcial

- La información entregada por PYGAM nos permite identificar la forma de cada función de identidad a nivel de atributo.
- Informa sobre la cantidad de splines, el lambda afectado y el nivel de significancia asociado.
- El problema es que la función de identidad no puede ser interpretada como el efecto lineal en la mayoría de los casos (bajo ciertas condiciones esto es posible).
- Para ello implementamos métodos de dependencia parcial (Friedman, 2001)
- Evalúa el efecto marginal de un atributo (o función de suavización específica) en los puntajes predichos del modelo.
- Permite resumir de manera simple (de manera relativamente no paramétrica) cuál es el comportamiento de un atributo específico en el vector objetivo.

Dependencia Parcial

$$\hat{f}_{x_s}(x_s) = \mathbb{E}_{x_c} \left[\hat{f}(x_s, x_c) \right]$$

- Para evaluar un atributo específico, evaluamos su función de suavización.

Dependencia Parcial

$$\hat{f}_{x_s}(x_s) = \mathbb{E}_{x_c} \left[\hat{f}(x_s, x_c) \right]$$

- Para evaluar un atributo específico, evaluamos su función de suavización.
- Esta función de suavización va a corresponder a la **esperanza matemática de la función**.

Dependencia Parcial

$$\hat{f}_{x_s}(x_s) = \mathbb{E}_{x_c} \left[\hat{f}(x_s, x_c) \right]$$

- Para evaluar un atributo específico, evaluamos su función de suavización.
- Esta función de suavización va a corresponder a la **esperanza matemática de la función**.
- Esta función depende de la **marginalización de todos los demás atributos** en nuestro modelo.

Dependencia Parcial

$$\hat{f}_{x_s}(x_s) = \mathbb{E}_{x_c} \left[\hat{f}(x_s, x_c) \right]$$

- Para evaluar un atributo específico, evaluamos su función de suavización.
- Esta función de suavización va a corresponder a la **esperanza matemática de la función**.
- Esta función depende de la **marginalización de todos los demás atributos** en nuestro modelo.
- Con lo cual podemos evaluar el comportamiento de nuestro atributo específico **controlando por todos los demás factores**.

{desafío}
latam_

*Academia de
talentos digitales*

www.desafiolatam.com