Modelos de ensamble (parte III)

Clase sincrónica

{desafío} latam_

Implementar ensambles de modelos en problemas complejos, ajustando diferentes factores para optimizar la predicción.

- **Unidad 1: Modelos de ensamble**
  (Parte I)

  (Parte II)

  (Parte III)

  Te encuentras aquí
- **Unidad 2: Redes neuronales**
  (Parte I)

  (Parte II)
- **Unidad 3: Procesamiento y Redes recurrentes**
  (Parte I)

  (Parte II)

{desafío} latam_

¿Qué aprenderás en esta sesión?

Comprender el funcionamiento de los principales modelos de regularización: Ridge, Lasso y Elastic Net y su aplicación.

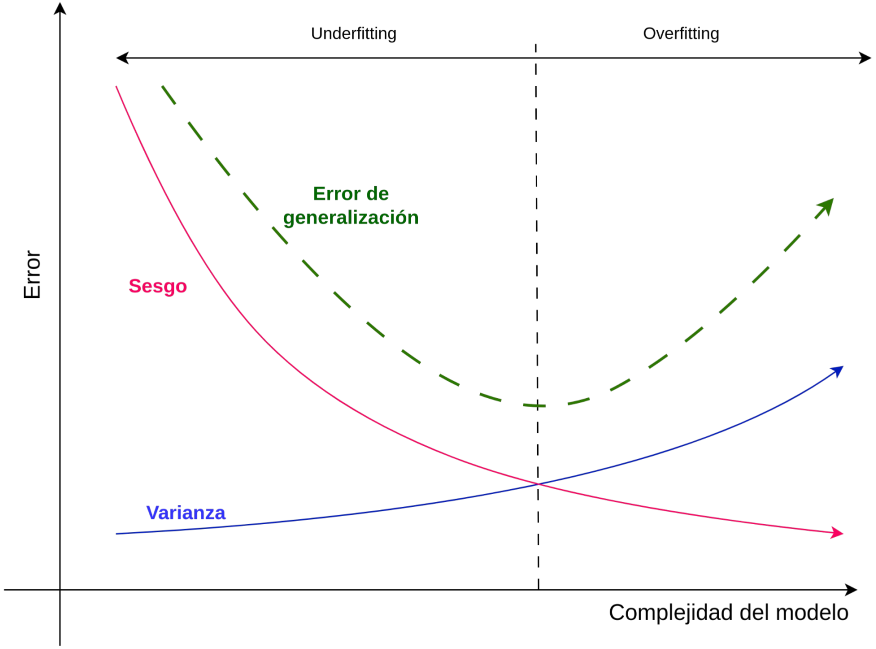
{desafío}
latam_

¿ Cúal es la diferencia entre sobre aprendizaje y sub aprendizaje ?

/* Sesgo-Varianza */

Sesgo-Varianza
Conceptos

- **Sesgo**: simplicidad excesiva de un modelo. Equivocada relación entre variables regresoras y target.
- **Varianza**: Sensibilidad de un modelo a pequeñas variaciones en los datos de entrenamiento. “Aprende ruido”
- **Trade-off**: Equilibrio sesgo-varianza. Punto en el cual se optimiza el nivel de error.



Sesgo-Varianza

Estrategias para controlar trade-off

- **Regularización**: permite reducir la varianza por medio de la penalización de la complejidad del modelo. Se logra así mejorar el nivel de generalización de los modelos.

- **Selección de características**: puede ayudar en la disminución del sesgo, al mejorar la representación del modelo.

- **Validación cruzada**: permite medir correctamente el rendimiento de los modelos y ajustar el sesgo-varianza

/* Regularización */

{desafío}
latam_

Regularización

Motivos para su uso

1 **Sobreajuste**

La regularización penaliza la complejidad de los modelos con esto se logra mejorar el nivel de generalización de estos.

2 **Multicolinealidad**

Cuando tenemos dos o más variables que explican lo mismo, genera problemas en ciertos modelo. Con la regularización pueden ser descartadas.

3 **Selección de características**

Con regularización podemos seleccionar características importantes.

4 **Reducción del ruido**

La regularización puede ayudar a reducir aquellas observaciones con mediciones extremas pudiendo ser errores medición

{desafío} latam_

Regularización

Regresión Lineal

$$\hat{\beta}_{MCO} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

En regresión lineal, para encontrar la mejor recta que se ajuste a los datos aplicamos mínimos cuadrados ordinarios (MCO), lo que corresponde a minimizar la diferencia cuadrática entre cada valor real y el valor predicho.

Sobre esta función explicaremos los diferentes modelos de regularización

Regularización Ridge

Mínimos cuadrados y sobreajuste

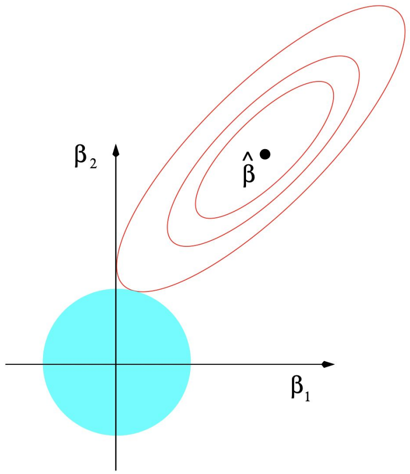
$$\hat{\beta}_{MCO} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

MCO en regresión lineal puede generar coeficientes (betas) con valores extremos que se distancian del resto. En consecuencia, se puede producir sobre-ajuste y perder capacidad de generalización.

Regularización Ridge
Norma L2

L2 penaliza los pesos o coeficientes asociados a las características, por medio de elevar al cuadrado cada peso.

$$L2 : ||\mathbf{w}||_2^2 = \sum_{j=1}^{m} w_j^2$$



Regularización Ridge

Hiper parámetro λ

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} \beta_i^2$$

- El hiper parámetro lambda se usa para controlar el nivel de penalización que se desea aplicar al modelo.
- Lambda domina la superficie de penalización, la cual está determinada por la cantidad de parámetros a inferir.
- Debido a la norma L2 (cuadrática) es que Ridge no puede descartar características de escasa importancia, sólo estará suavizando los coeficientes.

Sintonizando el hiper parámetro lambda
Validación cruzada

La optimización de nuestro modelo implica la elección cuidadosa de hiperparámetros. Para determinar el valor óptimo de lambda:

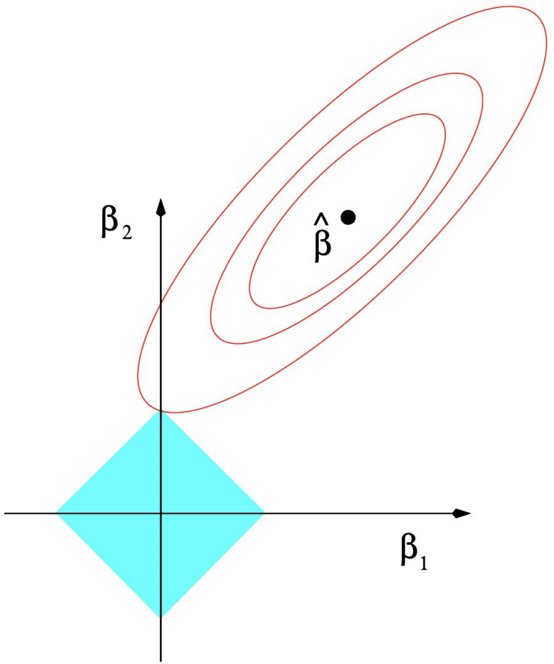
- definimos un conjunto de valores a probar.
- empleamos validación cruzada k-fold en el conjunto de entrenamiento para calcular el Error Cuadrático Medio (MSE) asociado con cada lambda.
- La selección del lambda óptimo se basa en la identificación del valor que minimiza el MSE en este proceso de validación cruzada

Regularización Lasso
Norma L1

Penaliza los pesos o coeficientes asociados a las características por medio del valor absoluto de cada peso.

$$L1 : ||\mathbf{w}||_1 = \sum_{j=1}^{m} |w_j|$$

Lasso: Least Absolute Shrinkage and Selection Operator.



Lasso

Hiper parámetro lambda

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} |\beta_i|$$

- El objetivo del hiper parámetro lambda es la misma que en Ridge, gobierna la superficie de penalización.
- A diferencia de Ridge, Lasso permite llevar a cero los coeficientes, con lo que logra descartar características.
- Lasso genera modelos "sparse" con muchos coeficientes cero, lo que es beneficioso en modelos con gran cantidad de atributos ya que reduce la complejidad del modelo.

Elastic Net

Hiperparámetros

$$\hat{\beta}^{enet} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i=1}^{p} \beta_i^2$$

- Elastic Net es una combinación de L1 y L2.
- Incluye dos hiperparámetros de regularización (ƛ1 y ƛ2), los cuales se pueden emplear para controlar el trade-off (sesgo-varianza).
- Aprovecha la característica de Lasso en cuanto a la selección de características, y con la parte de Ridge realizar una penalización parsimoniosa.

Conclusiones

Diferentes regularizaciones

¿Cuándo aplicar cada regularización?

¿Qué queremos lograr?

¿Qué efectos tiene una buena elección?

Actividad guiada "Abalone"

{desafío} latam_

Actividad guiada

Abalone

Veremos a continuación cómo aplicar estos modelos en Jupyter. Para esto, abre el archivo 01 - Implementación de normas y sigue las instrucciones que te dará tu profesor.

Desafío
“Enfermedad en la sangre”

Desafío

“Enfermedad en la sangre”

- Descarga el archivo “Desafío”.
- Tiempo de desarrollo asincrónico: desde 2 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪

Ideas fuerza

Los problemas de ***Sub aprendizaje y Sobre aprendizaje*** *están asociados a mejorar el trade-off entre sesgo y varianza.*

Los métodos de regularización: ***Ridge, Lasso y Elastic Net*** *se usan para controlar los problemas de sobre aprendizaje*

Ridge *suaviza los coeficientes sin la capacidad de selección de atributos.* ***Lasso*** *permite la selección de atributos.* ***Elastic Net*** *es una combinación entre Ridge y Lasso.*

Ajusta tu modelo: equilibra complejidad para predicciones más precisas y estables.

Próxima sesión...

Entender los componentes de una red neuronal feedforward.

{desafío} latam_

{desafío} latam_

Academia de talentos digitales