



Modelos de ensamble (parte III)

Tutoría

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



Autoaprendizaje

Recursos asincrónicos

- ¿Revisaste los recursos de la semana 3 (Guía y desafío)?
- ¿Tienes dudas sobre alguno de ellos?



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



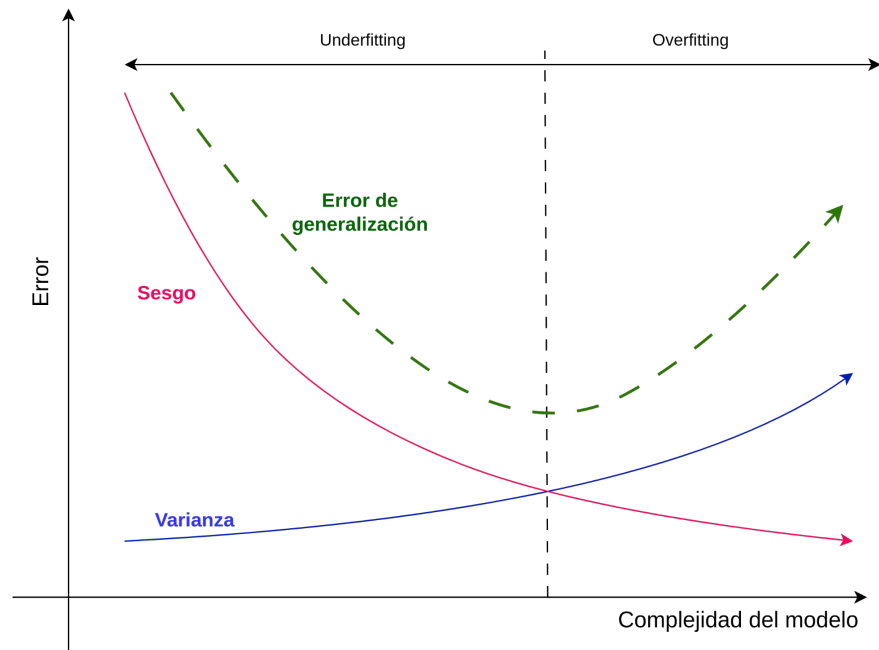
Ridge suaviza los coeficientes sin la capacidad de selección de atributos. **Lasso** permite la selección de atributos. **Elastic Net** es una combinación entre Ridge y Lasso.

/* 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 */

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 modelos. 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 de medición.

Regularizaciones

Ridge - Lasso

Ridge

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

Lasso

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

- El hiper parámetro lambda controla el nivel de penalización que se desea aplicar al modelo.
- Ridge no puede descartar características de escasa importancia, sólo estará suavizando los coeficientes.
- Lasso sí permite descartar características, por la norma que utiliza

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 (λ_1 y λ_2), 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"



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"

- ¿Leíste el desafío de esta semana? ¿Comprendes bien lo que se solicita en cada caso?
- ¿Hay contenidos que necesitas repasar antes de comenzar este desafío?
- ¿Necesitas algún ejemplo o indicación para alguna pregunta o requerimiento específico?



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



{desafío}
latam_

*Academia de
talentos digitales*

