



"A universe full of things, pixel art"

Temas extras

¿Qué es el sesgo en el aprendizaje automático?

El sesgo es un fenómeno que sesga el resultado de un algoritmo a favor o en contra de una idea.

El sesgo se considera un error sistemático que se produce en el propio modelo de aprendizaje automático debido a suposiciones incorrectas en el proceso de ML.

Técnicamente, podemos definir el sesgo como el error entre la predicción media del modelo y la realidad. Además, describe hasta qué punto el modelo se ajusta al conjunto de datos de entrenamiento:

- Un modelo con un sesgo más alto no se ajustará al conjunto de datos.
- Un modelo con un sesgo bajo se ajustará al conjunto de datos de entrenamiento.

¿Qué es el sesgo en el aprendizaje automático?

Las características de un modelo con un sesgo alto son las siguientes:

- No capta las tendencias adecuadas de los datos
- Posibilidad de underfitting
- Muy generalizado/excesivamente simplificado
- Tasa de error elevada

¿Qué es la varianza en el aprendizaje automático?

La varianza se refiere a los cambios en el modelo cuando se utilizan diferentes porciones del conjunto de datos de entrenamiento.

En pocas palabras, la varianza es la variabilidad en la predicción del modelo, es decir, cuánto puede ajustarse la función ML dependiendo del conjunto de datos dado. La varianza procede de modelos muy complejos con un gran número de características (features).

- Los modelos con un sesgo elevado tendrán una varianza baja.
- Los modelos con alta varianza tendrán un sesgo bajo.

Todo ello contribuye a la flexibilidad del modelo. Por ejemplo, un modelo que no se ajusta a un conjunto de datos con un sesgo alto creará un modelo inflexible con una varianza baja que da lugar a un modelo de aprendizaje automático subóptimo.

¿Qué es la varianza en el aprendizaje automático?

Las características de un modelo de alta varianza incluyen:

- Ruido en el conjunto de datos
- Posibilidad de sobreajuste
- Modelos complejos
- Intentar poner todos los puntos de datos lo más cerca posible

Underfitting & overfitting

Los términos underfitting y overfitting se refieren a la forma en que el modelo no se ajusta a los datos. El ajuste de un modelo está directamente relacionado con la precisión de las predicciones de un conjunto de datos determinado.

- El underfitting se produce cuando el modelo es incapaz de hacer coincidir los datos de entrada con los datos a estimar. Esto ocurre cuando el modelo no es lo suficientemente complejo como para ajustarse a todos los datos disponibles y tiene un rendimiento pobre con el conjunto de datos de entrenamiento.
- El overfitting se refiere a los casos en los que el modelo intenta ajustarse a datos inexistentes. Esto ocurre cuando se trata de modelos muy complejos en los que el modelo se ajusta a casi todos los puntos de datos dados y tiene un buen rendimiento en los conjuntos de datos de entrenamiento. Sin embargo, el modelo no sería capaz de generalizar el punto de datos en el conjunto de datos de prueba para predecir el resultado con precisión.

Bias vs variance: A trade-off

El sesgo y la varianza están relacionados de forma inversa. Es imposible tener un modelo ML con un sesgo bajo y una varianza baja.

Cuando uno modifica un algoritmo de ML para que se ajuste mejor a un conjunto de datos determinado, conseguirá un sesgo bajo, pero aumentará la varianza. De este modo, el modelo se ajustará al conjunto de datos, pero aumentará las posibilidades de predicciones inexactas.

Lo mismo ocurre cuando se crea un modelo de baja varianza con un sesgo más alto. Aunque reducirá el riesgo de predicciones inexactas, el modelo no se ajustará correctamente al conjunto de datos.

Se trata de un delicado equilibrio entre el sesgo y la varianza. Sin embargo, es importante señalar que tener una varianza más alta no indica que el algoritmo de ML sea malo. Los algoritmos de aprendizaje automático deberían ser capaces de manejar cierta varianza.

Podemos abordar el equilibrio de múltiples maneras...

Bias vs variance: A trade-off

Aumentar la complejidad del modelo para tener en cuenta el sesgo y la varianza, disminuyendo así el sesgo general y aumentando la varianza hasta un nivel aceptable. De este modo, el modelo se ajusta al conjunto de datos de entrenamiento sin incurrir en errores de varianza significativos.

Aumentar el conjunto de datos de entrenamiento también puede ayudar a equilibrar esta compensación, hasta cierto punto. Este es el método preferido cuando se trata de modelos con sobreajuste. Además, esto permite a los usuarios aumentar la complejidad sin que los errores de varianza contaminen el modelo, como ocurre con un conjunto de datos grande.

Un conjunto de datos grande ofrece más puntos de datos para que el algoritmo pueda generalizar los datos fácilmente. Sin embargo, el principal problema de aumentar el conjunto de datos de entrenamiento es que el underfitting o bajo sesgo no son tan sensibles al conjunto de datos de entrenamiento. Por lo tanto, aumentar los datos es la solución preferida cuando se trata de modelos de alta varianza y alto sesgo.

the bias vs. variance trade-off

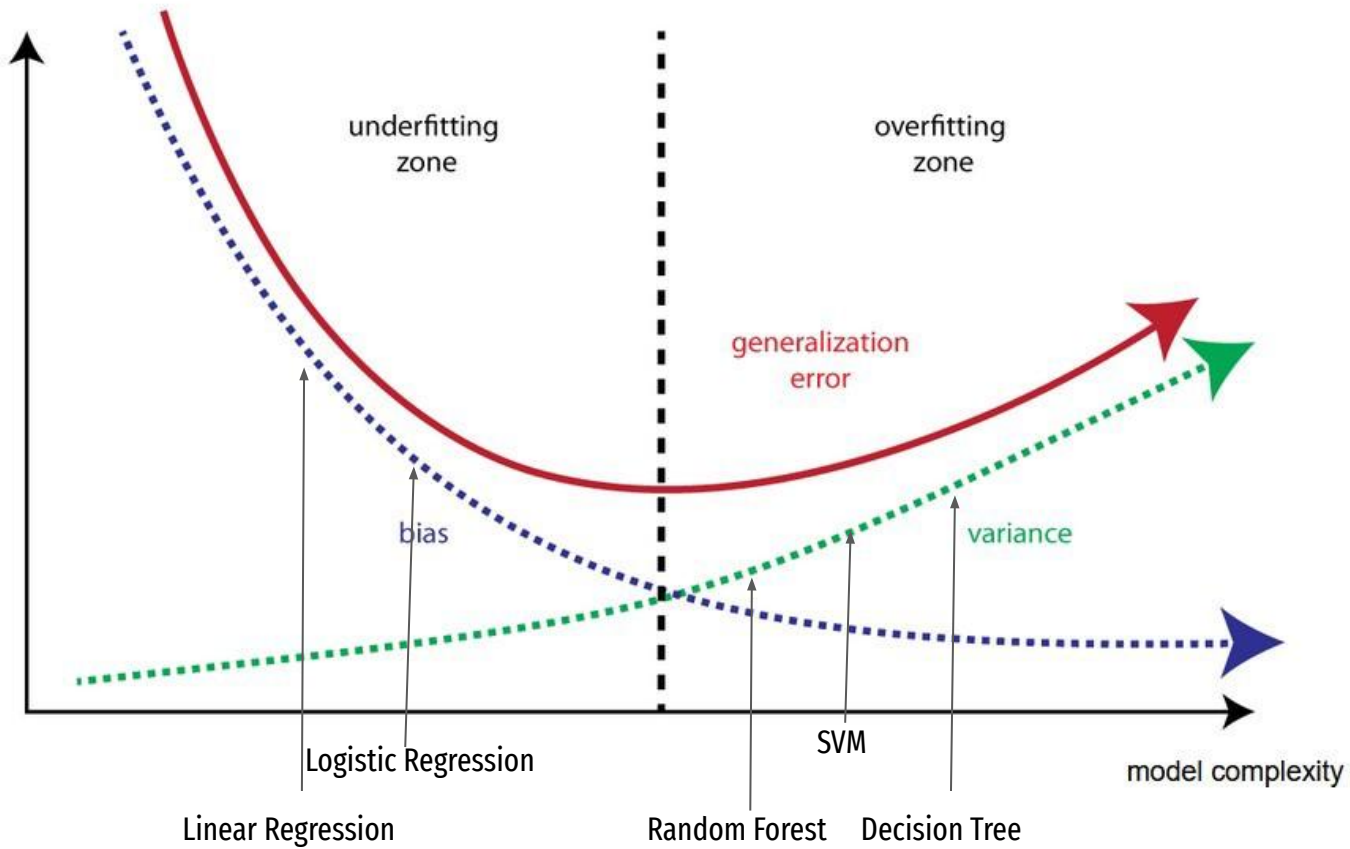


Gráfico esquemático

Es fundamental tener en cuenta el sesgo y la varianza

El sesgo y la varianza son dos componentes clave que hay que tener en cuenta al desarrollar cualquier modelo de aprendizaje automático bueno y preciso.

- **El sesgo crea errores consistentes** en el modelo de ML, lo que representa un modelo de ML más simple que no es adecuado para un requisito específico.
- Por otro lado, **la varianza crea errores de varianza** que conducen a predicciones incorrectas viendo tendencias o puntos de datos que no existen.

Como profesional hay que tener en cuenta estos dos factores al crear un modelo ML. Por lo general, el objetivo es mantener el sesgo lo más bajo posible mientras se introducen niveles aceptables de varianza. Esto puede hacerse aumentando la complejidad o incrementando el conjunto de datos de entrenamiento.

La idea es de forma equilibrada crear un modelo de aprendizaje automático aceptable.



"A linear regression, digital art"

RL extras

¿Cómo se hace una regresión lineal ordinaria?

$$\hat{y} = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

Donde p es el número de features, w junto con b son los parámetros a aprender por el modelo.

La manera más simple y clásica para obtener los parámetros b y w es minimizar el error cuadrático medio entre las predicciones y los targets reales sobre el conjunto de entrenamiento.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

¿Se puede tener overfitting con una RL?

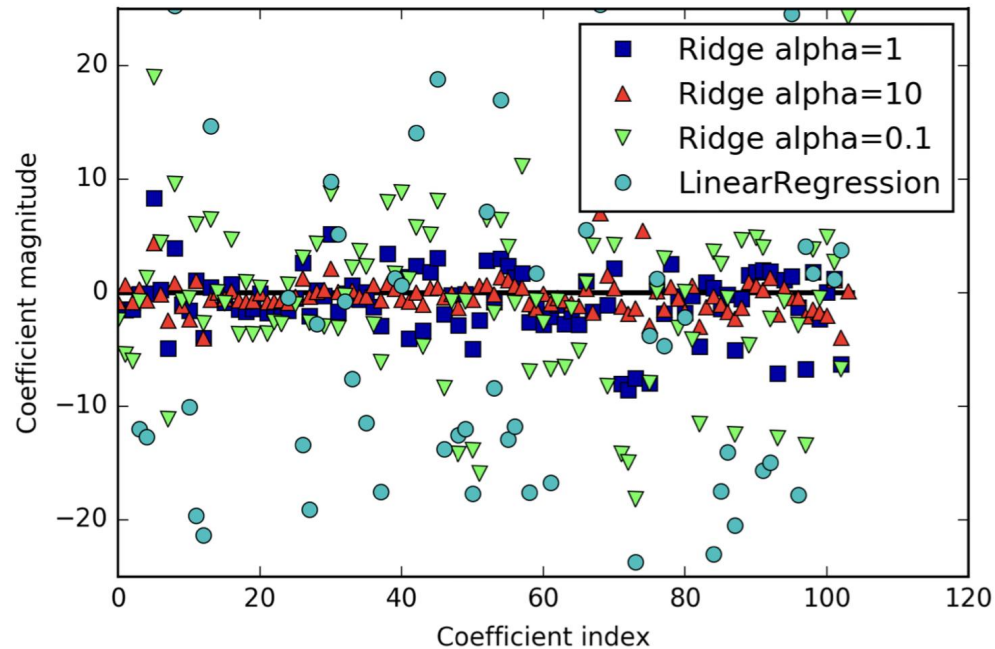
Regresión Ridge

$$\hat{y} = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

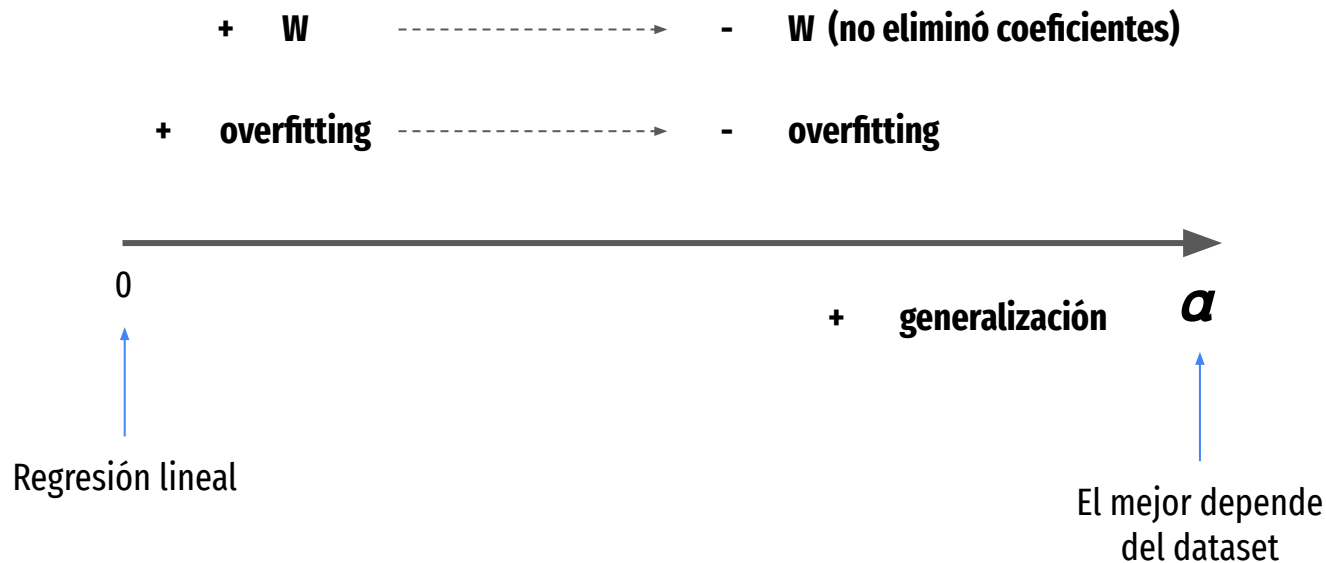
Regularización \longrightarrow **Evitar overfitting** \longrightarrow **L2**

$$\text{Cost} = \underbrace{\sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2}_{\text{Loss function}} + \underbrace{\lambda \sum_{j=0}^M W_j^2}_{\text{Regularization Term}}$$

Regresión Ridge



Regresión Ridge



Regresión Lasso

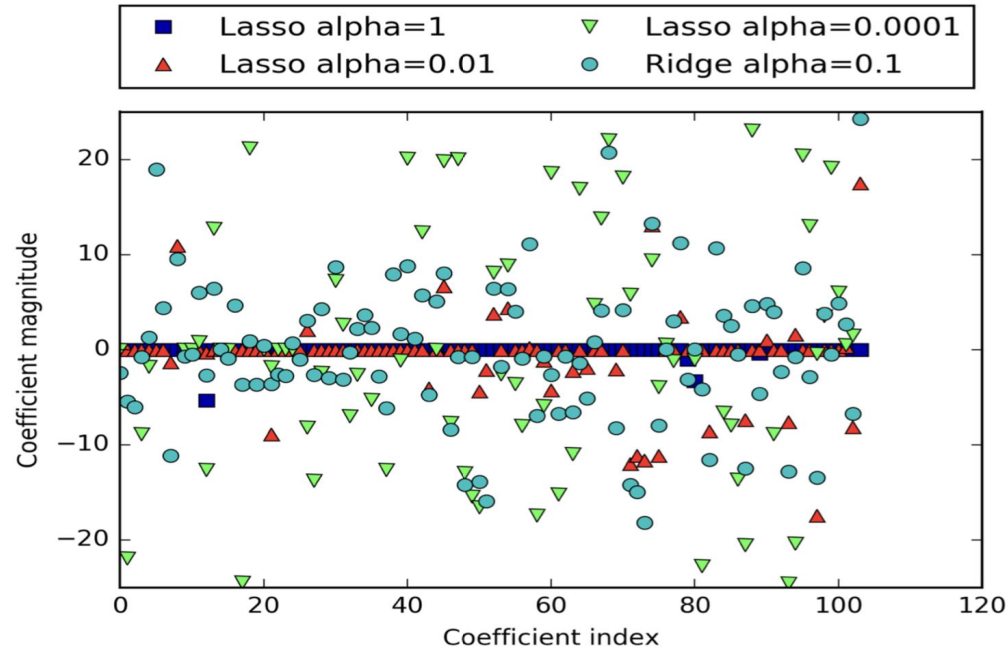
$$\hat{y} = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

Regularización \longrightarrow **Evitar overfitting + EF** \longrightarrow **L1**

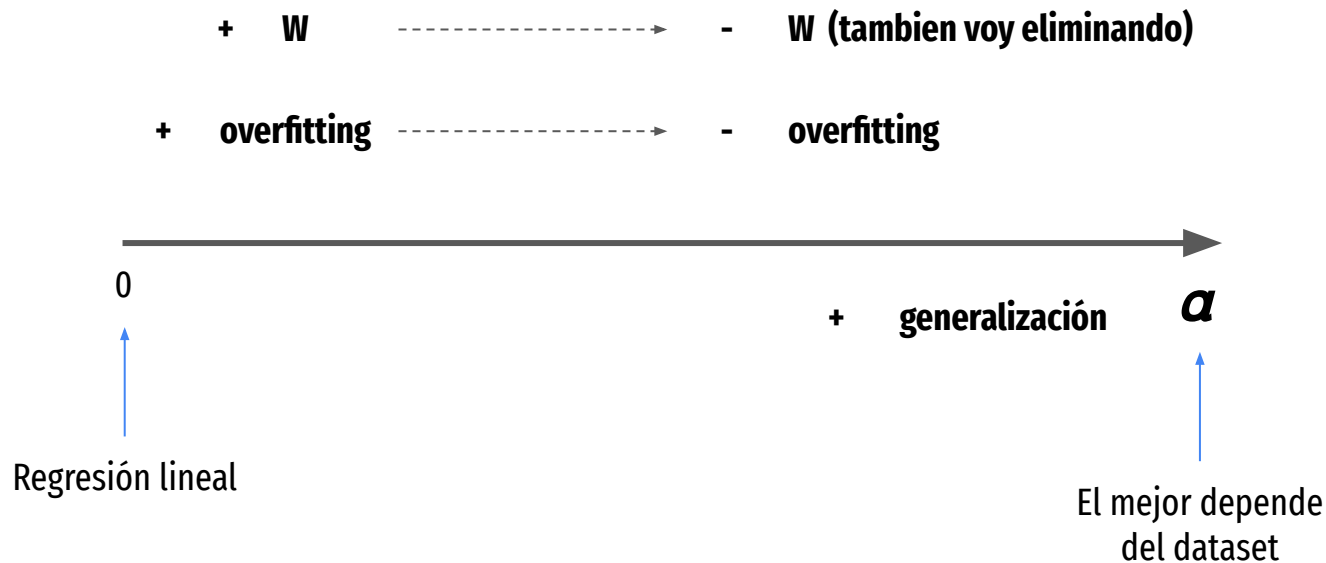
$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \alpha \sum_{j=0}^M |W_j|$$

El parámetro alpha controla cuán fuertemente los coeficientes son enviados a cero.

Regresión Lasso



Regresión Lasso



ElasticNet

Regresión lineal con una combinación de normas L1 y L2 como regularizadores.

$$\text{minimize: } RSS + Lasso + Ridge = \sum_{i=1}^n \left(y_i - \left(\beta_0 + \sum_{j=1}^p \beta_j x_{ij} \right) \right)^2 + \alpha_L \sum_{j=1}^p |\beta_j| + \alpha_R \sum_{j=1}^p \beta_j^2$$