

# IC-6200 Inteligencia Artificial

## Apuntes del 27 de Febrero, 2024

### - Sesgo y Varianza -

Ronaldo Vindas B. 2018238697

#### RESPUESTAS DEL QUIZ #2

- 1) Mencione dos razones del por qué es conveniente escoger MSE en lugar de MAE como loss function.

Respuesta:

- MSE es derivable y permite usar algoritmos como descenso gradiente.
- Garantiza la existencia de un Mín. Global.

- 2) Mencione qué es un mínimo local y un mínimo global en una función (20 pts).

Respuesta:

- Un mínimo local de una función es un punto en el que el valor de la función es menor que en puntos cercanos, pero posiblemente mayor que en un punto alejado.
- Un mínimo global es un punto en el que el valor de la función es menor que en otros puntos factibles.

- 3) Desarrolle la derivada parcial de L con respecto a W de:

$$\frac{1}{N} \sum ((wx_i + b) - y_i)^2$$

(35 pts).

Respuesta:

$$\frac{\partial L}{\partial w} = \frac{1}{N} \sum 2((wx_i + b) - y_i) \cdot x_i$$

- 4) Desarrolle la derivada parcial de L con respecto a b de

$$\frac{1}{N} \sum ((wx_i + b) - y_i)^2$$

(35 pts).

Respuesta:

$$\frac{\partial L}{\partial w} = \frac{1}{N} \sum 2((wx_i + b) - y_i)$$

- 5) ¿El Learning Rate del Descenso Gradiente es un parámetro o hiper-parámetro? (5 pts).

Respuesta:

- Es un hiper-parámetro que corresponde al alfa, lo selecciona el ingeniero o Data Scientist que está trabajando el algoritmo.

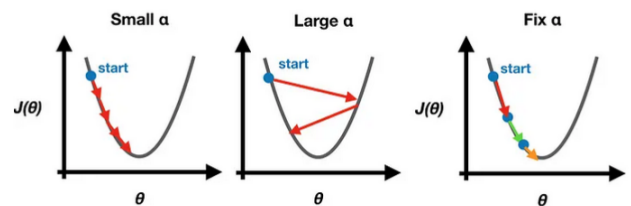
#### - REPASO DE REGRESIÓN LINEAL -

Se debe optimizar lo mayor posible la función de pérdida para poder realizar una actualización de parámetros:

$$w = w - \alpha \frac{\partial L}{\partial w}$$

$$b = b - \alpha \frac{\partial L}{\partial b}$$

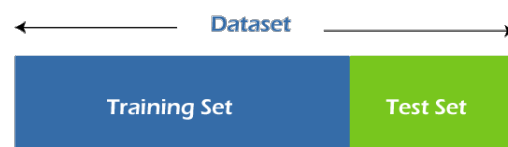
- "α" (alpha) es un hiper-parámetro.
- Indica con que "fuerza" nos movemos a través del Gradiente Descendiente.



- Un alpha muy **pequeño** es lento y va a requerir muchas actualizaciones para poder llegar a converger en una solución óptima.
- Un alpha muy **grande**, al hacer actualizaciones drásticas, puede provocar comportamientos muy divergentes.
- Un alpha **adecuado** puede llegar a alcanzar rápidamente el punto mínimo. Se puede averiguar haciendo experimentos corriendo el modelo varias veces con diferentes alpha's.[1]

#### - SESGO Y VARIANZA -

Conviene cortar el dataset en dos partes: **Training (80%)** y **Testing (20%)**. La idea es poder aplicar esto tanto en las tareas como en proyectos.

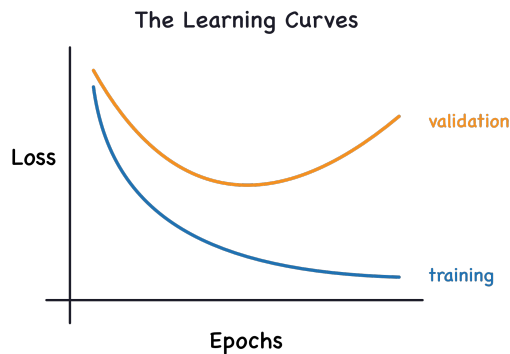


- **Training Dataset:** Es usado para ajustar el modelo. El modelo identifica patrones. Relaciones entre variables. Debería representar la diversidad, de escenarios que se espera encontrar.

- **Testing Set:** Es usado para evaluar el modelo con ejemplos que NO utilizó durante el entrenamiento. Debe ser independiente del set de entrenamiento. Se utilizan métricas para estimar el rendimiento del modelo (Accuracy, Matrices de Confusión, F1-Score, etc).

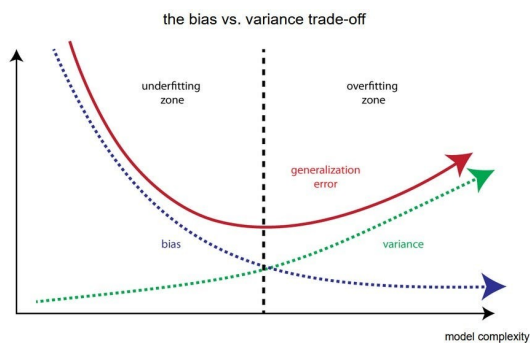
### Posibles Escenarios en Validación y Entrenamiento:

- 1) **Bajo error en el Training - Bajo error en el Testing:**  
Es el escenario ideal. El modelo evita el ruido existente en los datos y logró aprender y puede generalizar correctamente. Puede ver bien los demás ejemplos.
- 2) **Bajo error en el Training - Alto Error en el Testing:**  
Ocurre un **Overfitting**. Quiere decir que el modelo aprendió mucho de los datos, tanto así que ya no es capaz de generalizar. Tiene una varianza alta entre los datos que se aprendieron y se predicen.
- 3) **Alto error en el Training - Alto error en el Testing:**  
Ocurre un **Underfitting**. Significa que el modelo es demasiado simple y que no está aprendiendo nada de los datos. La información que tiene es insuficiente. Tiene un alto sesgo.[2]



### Bias / Variance Tradeoff:

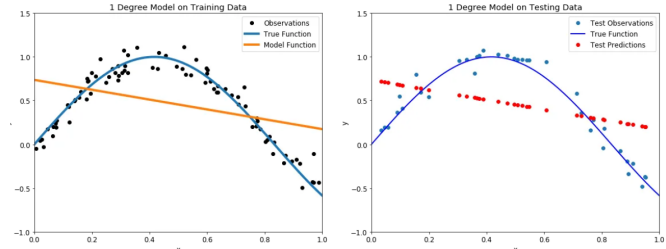
Entonces, necesitamos que nuestro modelo tenga una baja varianza y un bajo sesgo, es decir, optimizar el error en el modelo y mantenerlo lo más bajo posible.



### ¿Qué representa tener un Alto Bias (Sesgo)?

Significa que el modelo comete muchos errores en el Training Set. El modelo es simple y asume mucho del Training Set y No usa todos los features del modelo (**Underfitting**).

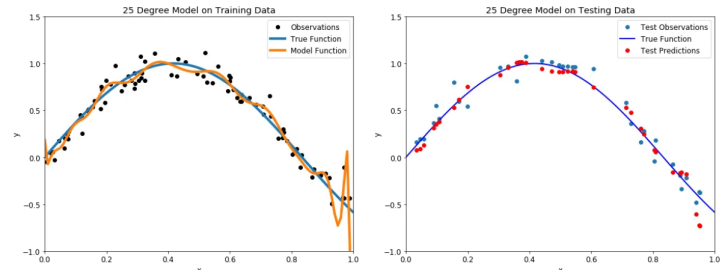
Se puede evitar usando un modelo más complejo (Regresión Polinomial, Redes Neuronales, etc), evitar usar features del Training Set que no son adecuados para el problema y que no tienen la capacidad para hacer predicciones. [3]



### ¿Qué representa tener una Alta Varianza?

Significa que el modelo es muy complejo y se ajusta mucho a los datos de entrenamiento, no es capaz de generalizar y es sensible a las variaciones de los datos. Suele suceder con datos de alta dimensionalidad y pocos ejemplos (**Overfitting**).

Se puede evitar usando un modelo que sea más simple, reduciendo la dimensionalidad, obteniendo más ejemplos en el Training Set (Aunque no siempre es posible) y aplicar técnicas de regularización que hacen el modelo más simple. [3]



### REFERENCES

- [1] J. Jordan, "Setting the learning rate of your neural network." Jeremy Jordan, 03 2018. [Online]. Available: <https://www.jeremyjordan.me/nn-learning-rate/>
- [2] W. Koehrsen, "Overfitting vs. underfitting: A complete example - towards data science," Medium, 01 2018. [Online]. Available: <https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765>
- [3] S. Singh, "Understanding the bias-variance tradeoff," Medium, 05 2018. [Online]. Available: <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>