

Aprendizaje automático

Clase 5

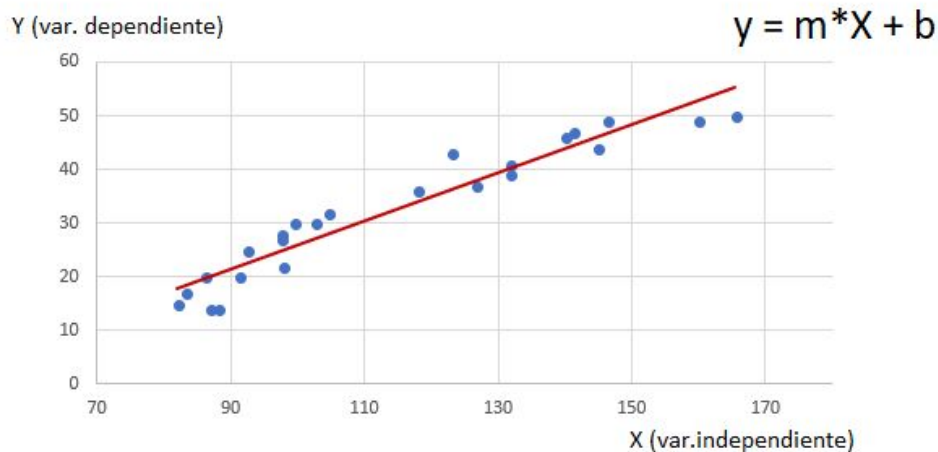
Martin Pustilnik, Iris Sattolo,
Maximiliano Beckel



Modelo de Regresión Lineal

- Como habíamos mencionado en clases anteriores, un modelo de regresión es aquel modelo que intenta predecir el valor de una variable numérica continua.
- En un Modelo de Regresión Lineal (MRL) se busca establecer una relación lineal entre la variable que se busca explicar/predecir (que llamaremos variable dependiente) y una o más variables explicativas/predictoras (que llamaremos variables independientes).
- Si en el modelo tenemos como una variable independiente es un modelo de regresión lineal univariado, pero también podemos tener varias variables independientes y, en tal caso, llamaremos al modelo regresión lineal multivariada

¿Qué significa una relación lineal?



m : pendiente (¿cuánto cambia y cuando cambia x?)
b : ordenada al origen (valor de y cuando x = 0)

Regresión Lineal Multivariada:

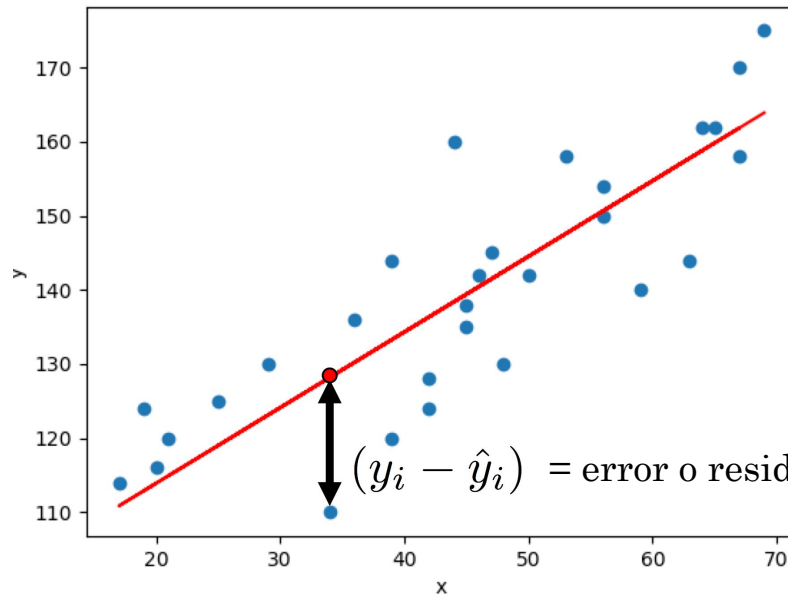
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$

X_1, X_2, \dots, X_i son distintas variables independientes

Modelo de Regresión Lineal

- Como habíamos mencionado en clases anteriores, un modelo de regresión es aquel modelo que intenta predecir el valor de una variable numérica continua.
- En un Modelo de Regresión Lineal (MRL) se busca establecer una relación lineal entre la variable que se busca explicar/predecir (que llamaremos variable dependiente) y una o más variables explicativas/predictoras (que llamaremos variables independientes).
- Si en el modelo tenemos como una variable independiente es un modelo de regresión lineal univariado, pero también podemos tener varias variables independientes y, en tal caso, llamaremos al modelo regresión lineal multivariada

¿Qué error tiene mi modelo?



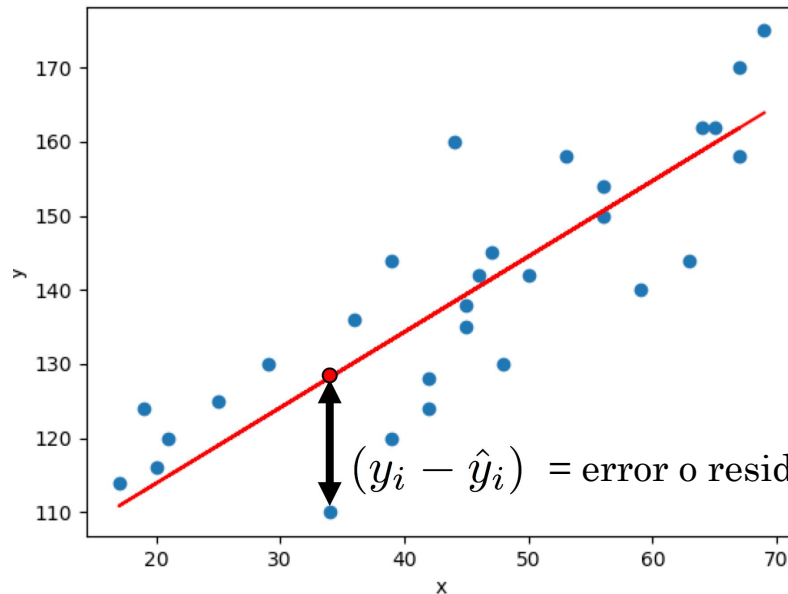
$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Componente Lineal del Modelo}} + \underbrace{\varepsilon_i}_{\text{Componente de error aleatorio}}$$

$(y_i - \hat{y}_i) = \text{error o residuo del modelo para } X_i$

Modelo de Regresión Lineal

- Como habíamos mencionado en clases anteriores, un modelo de regresión es aquel modelo que intenta predecir el valor de una variable numérica continua.
- En un Modelo de Regresión Lineal (MRL) se busca establecer una relación lineal entre la variable que se busca explicar/predecir (que llamaremos variable dependiente) y una o más variables explicativas/predictoras (que llamaremos variables independientes).
- Si en el modelo tenemos como una variable independiente es un modelo de regresión lineal univariado, pero también podemos tener varias variables independientes y, en tal caso, llamaremos al modelo regresión lineal multivariada

Cálculo del Error de mi Modelo



Error cuadrático medio:

N: número de
observaciones

Valor
observado

Valor
predicho

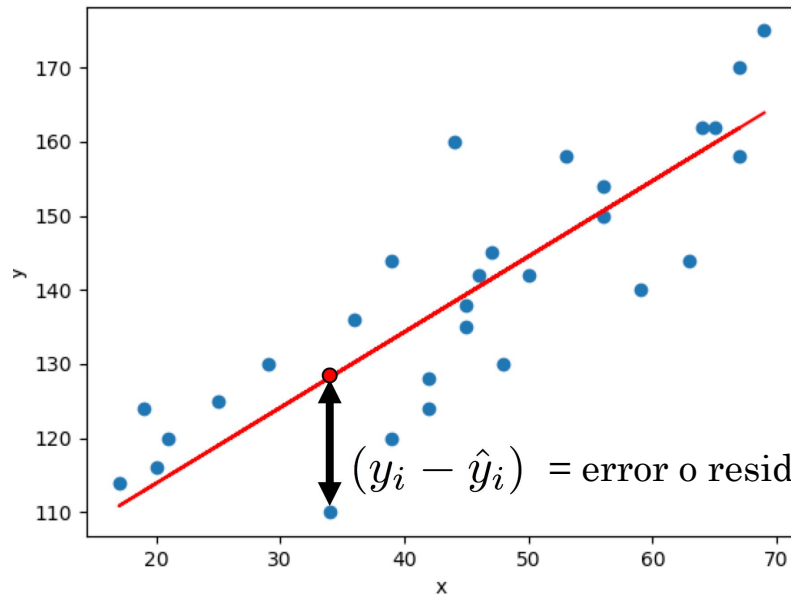
$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$(y_i - \hat{y}_i)$ = error o residuo del modelo para X_i

Modelo de Regresión Lineal

- Como habíamos mencionado en clases anteriores, un modelo de regresión es aquel modelo que intenta predecir el valor de una variable numérica continua.
- En un Modelo de Regresión Lineal (MRL) se busca establecer una relación lineal entre la variable que se busca explicar/predecir (que llamaremos variable dependiente) y una o más variables explicativas/predictoras (que llamaremos variables independientes).
- Si en el modelo tenemos como una variable independiente es un modelo de regresión lineal univariado, pero también podemos tener varias variables independientes y, en tal caso, llamaremos al modelo regresión lineal multivariada

Cálculo del Error de mi Modelo



Error cuadrático medio:

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

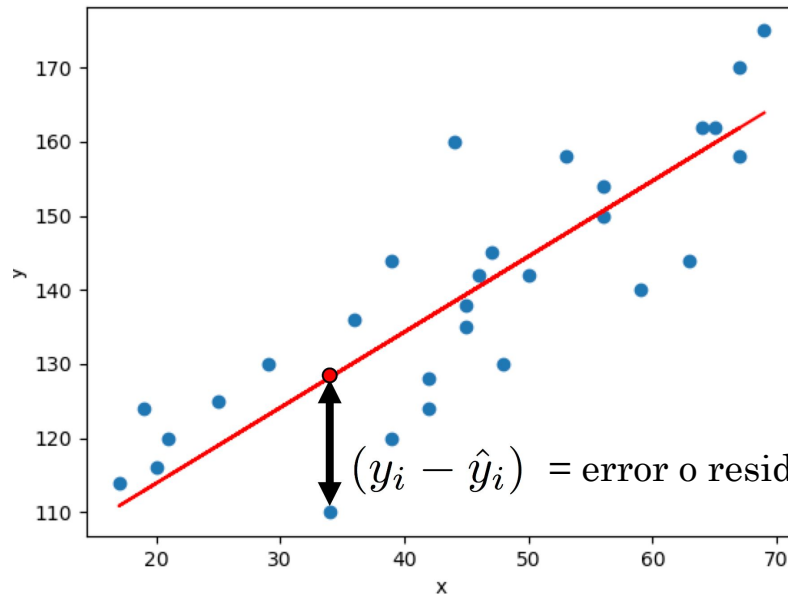
Al realizar un modelo de regresión lineal lo que buscamos es encontrar la ecuación lineal que minimice el error (es decir, la distancia de nuestros datos a la recta)

$(y_i - \hat{y}_i)$ = error o residuo del modelo para X_i

Modelo de Regresión Lineal

- Como habíamos mencionado en clases anteriores, un modelo de regresión es aquel modelo que intenta predecir el valor de una variable numérica continua.
- En un Modelo de Regresión Lineal (MRL) se busca establecer una relación lineal entre la variable que se busca explicar/predecir (que llamaremos variable dependiente) y una o más variables explicativas/predictoras (que llamaremos variables independientes).
- Si en el modelo tenemos como una variable independiente es un modelo de regresión lineal univariado, pero también podemos tener varias variables independientes y, en tal caso, llamaremos al modelo regresión lineal multivariada

Cálculo del Error de mi Modelo



Error cuadrático medio:

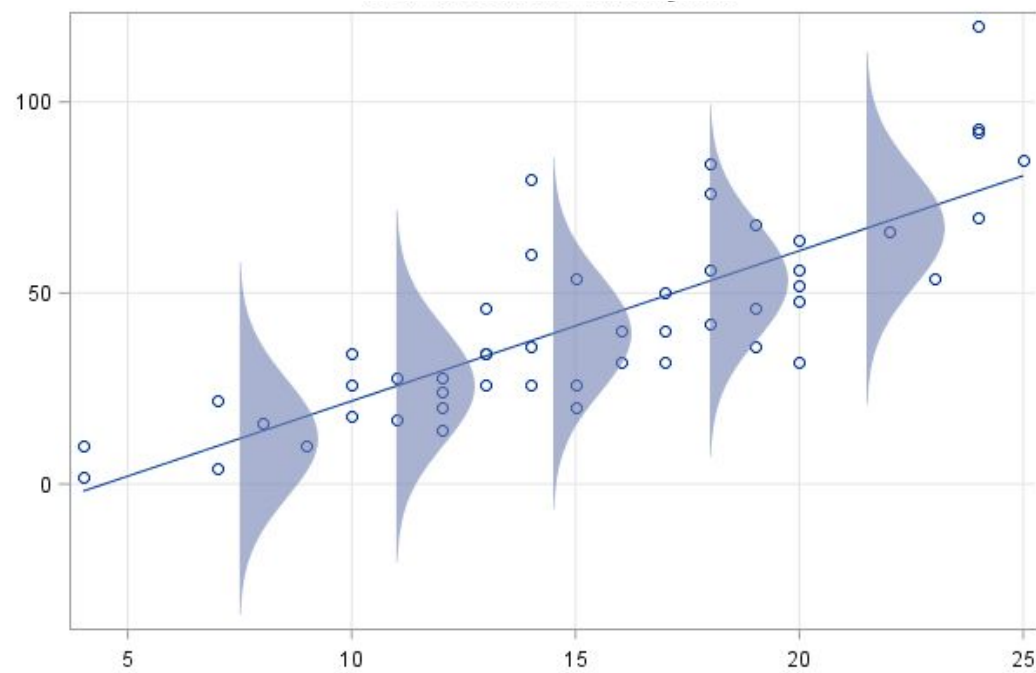
$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

En general, cualquier modelo de aprendizaje automático no es otra cosa que un proceso de optimización que busca minimizar una **función de costo**.

Modelo de Regresión Lineal

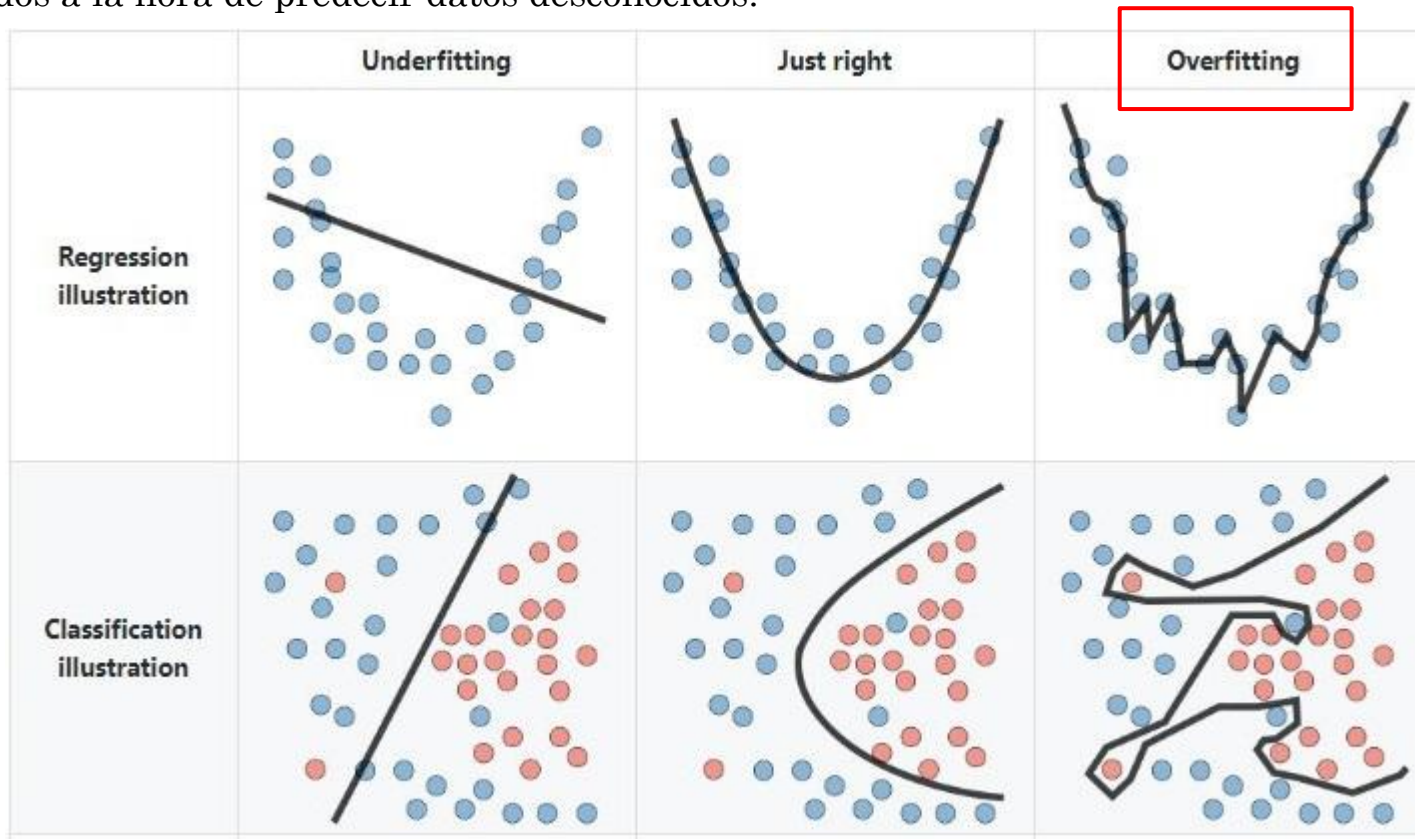
Supuestos:

- Linealidad: la variable dependiente y las variables independientes mantienen una relación lineal entre sí.
- Independencia: Las observaciones son independientes entre sí.
- Homocedasticidad: para todo el rango de las variables independientes, la varianza del error en mi modelo es constante.
- Normalidad: los errores en el modelo siguen una distribución normal
- No colinealidad: no existe una alta correlación entre las variables independientes.



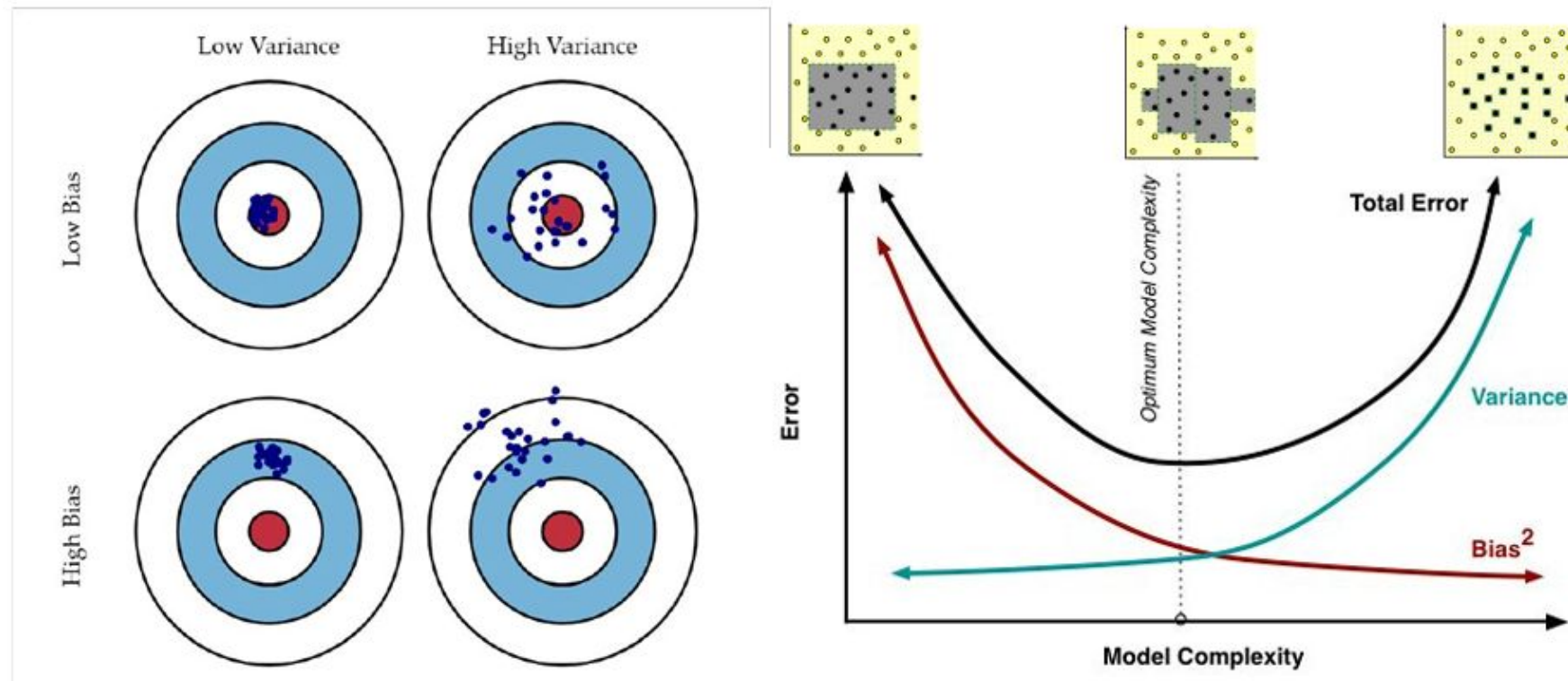
Problema: relación sesgo-varianza

Nuestra misión es construir un modelo que no solo logre representar bien los datos que tengo, sino también, que permita predecir nuevos datos que nunca vió (diferencia entre memorizar y aprender). Cuanto más complejo sea nuestro modelo (más parámetros tenga) posiblemente le vaya mejor explicando los datos que tengo pero le va a tener peores resultados a la hora de predecir datos desconocidos.



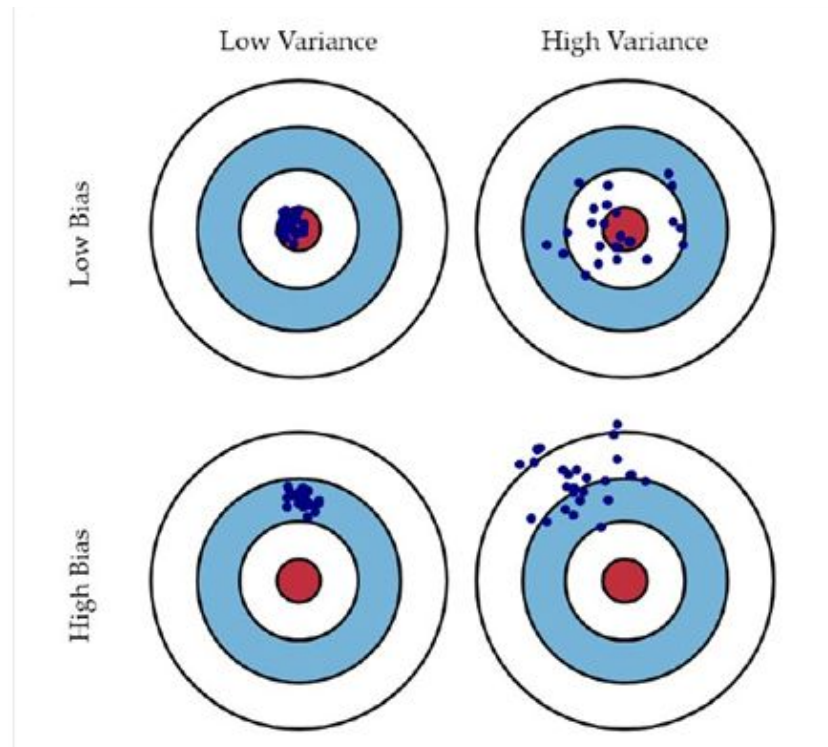
Problema: relación sesgo-varianza

Nuestra misión es construir un modelo que no solo logre representar bien los datos que tengo, sino también, que permita predecir nuevos datos que nunca vió (diferencia entre memorizar y aprender). Cuando más complejo sea nuestro modelo (más parámetros tenga) posiblemente le vaya mejor explicando los datos que tengo pero le va a tener peores resultados a la hora de predecir datos desconocidos.



Problema: relación sesgo-varianza

Nuestra misión es construir un modelo que no solo logre representar bien los datos que tengo, sino también, que permita predecir nuevos datos que nunca vió (diferencia entre memorizar y aprender). Cuanto más complejo sea nuestro modelo (más parámetros tenga) posiblemente le vaya mejor explicando los datos que tengo pero le va a tener peores resultados a la hora de predecir datos desconocidos.

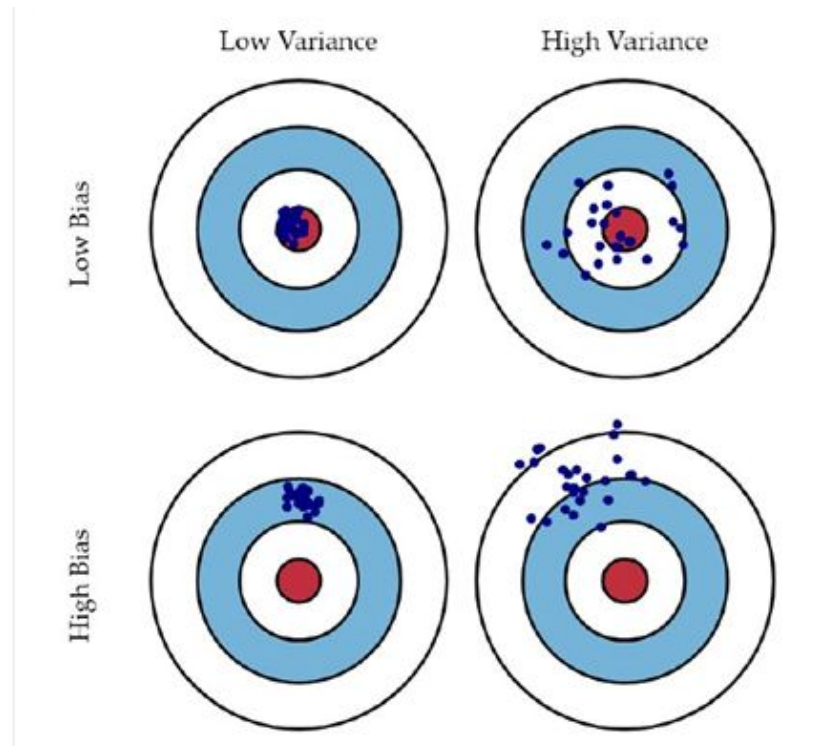


El **sesgo** o Bias en un modelo de machine learning es un tipo de error que indica la diferencia que existe entre la predicción del modelo y el valor actual. Si el modelo tiene un Bias alto significa que le presta poca atención a los datos y sobre simplifica el modelo.

La **varianza** también también es parte del error en un modelo de machine learning. Este error lo podemos entender como que tan sensible es nuestro modelo a los datos. Si tenemos exceso de sensibilidad el modelo puede creer ver patrones que realmente no existen ahí. Si tenemos un valor de varianza alto esto significa que el modelo le presta mucha atención a los datos de entrenamiento y no va a generalizar bien en datos que no ha visto.

Problema: relación sesgo-varianza

Nuestra misión es construir un modelo que no solo logre representar bien los datos que tengo, sino también, que permita predecir nuevos datos que nunca vió (diferencia entre memorizar y aprender). Cuanto más complejo sea nuestro modelo (más parámetros tenga) posiblemente le vaya mejor explicando los datos que tengo pero le va a tener peores resultados a la hora de predecir datos desconocidos.



El **sesgo** o Bias en un modelo de machine learning es un tipo de error que indica la diferencia que existe entre la predicción del modelo y el valor actual. Si el modelo tiene un Bias alto significa que le presta poca atención a los datos y sobre simplifica el modelo.

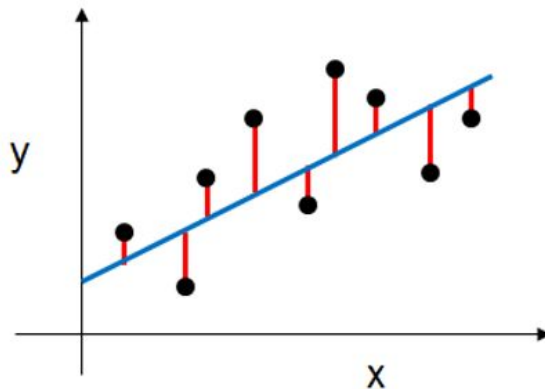
La **varianza** también también es parte del error en un modelo de machine learning. Este error lo podemos entender como que tan sensible es nuestro modelo a los datos. Si tenemos exceso de sensibilidad el modelo puede creer ver patrones que realmente no existen ahí. Si tenemos un valor de varianza alto esto significa que el modelo le presta mucha atención a los datos de entrenamiento y no va a generalizar bien en datos que no ha visto.

$$Error(x) = Bias^2 + Variance + ErrorIrreducible$$

Volvamos a MRL: Regularización

La **regularización** es una técnica que nos permite ajustar un modelo evitando el sobreajuste de nuestros datos (overfitting). ¿Cómo lo logra? Incluyendo en la función de costo una penalización que haga reducir el valor de los coeficientes de mi modelo. De esta forma, evito que mi modelo le dé “demasiada importancia” a los datos con los que entreno mi modelo, reduciendo la varianza y evitando así el overfitting. En esta técnica se agrega un hiperparametro (ya veremos la diferencia que tiene con un parámetro) que suele representarse con la letra griega alfa o lambda y va a permitirnos elegir cuánta importancia le damos a la regularización.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$



p is the number of features

β_j is a model coefficient

α is a tuning parameter: Increasing the α penalizes the coefficients and thus shrinks them.

$$SS_{residuals} = \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

Model Prediction
↓
Observed Result

La Regresión Lasso suma en la función de costo el valor absoluto de los coeficientes. Este tipo de regularización suele llevar a que algunos coeficientes se hagan igual a cero (feature selection)

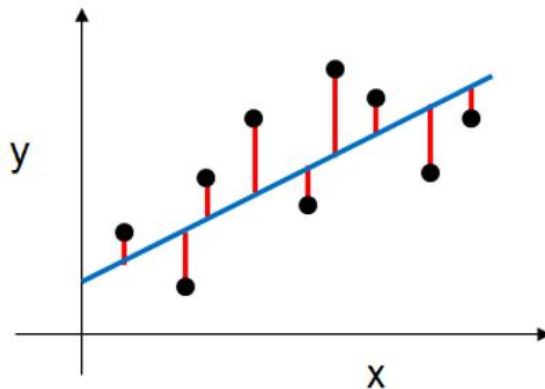
$$RSS + \alpha \sum_{j=1}^p |\beta_j|$$

Lasso regression
(or "L1 regularization") minimizes:

Volvamos a MRL: Regularización

La **regularización** es una técnica que nos permite ajustar un modelo evitando el sobreajuste de nuestros datos (overfitting). ¿Cómo lo logra? Incluyendo en la función de costo una penalización que haga reducir el valor de los coeficientes de mi modelo. De esta forma, evito que mi modelo le dé “demasiada importancia” a los datos con los que entreno mi modelo, reduciendo la varianza y evitando así el overfitting. En esta técnica se agrega un hiperparametro (ya veremos la diferencia que tiene con un parámetro) que suele representarse con la letra griega alfa o lambda y va a permitirnos elegir cuánta importancia le damos a la regularización.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$



p is the number of features

β_j is a model coefficient

α is a tuning parameter: Increasing the α penalizes the coefficients and thus shrinks them.

Ridge regression
(or "L2 regularization") minimizes:

$$RSS + \alpha \sum_{j=1}^p \beta_j^2$$

$$SS_{residuals} = \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

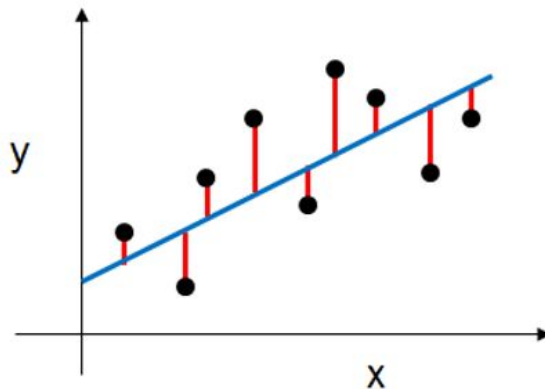
Model Prediction
↓
Observed Result

La Regresión Ridge suma en la función de costo el cuadrado de los coeficientes. Este tipo de regularización reduce el valor de los coeficientes pero no los llega a hacerlos igual a cero.

Volvamos a MRL: Regularización

La **regularización** es una técnica que nos permite ajustar un modelo evitando el sobreajuste de nuestros datos (overfitting). ¿Cómo lo logra? Incluyendo en la función de costo una penalización que haga reducir el valor de los coeficientes de mi modelo. De esta forma, evito que mi modelo le dé “demasiada importancia” a los datos con los que entreno mi modelo, reduciendo la varianza y evitando así el overfitting. En esta técnica se agrega un hiperparametro (ya veremos la diferencia que tiene con un parámetro) que suele representarse con la letra griega alfa o lambda y va a permitirnos elegir cuánta importancia le damos a la regularización.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$



p is the number of features

β_j is a model coefficient

α is a tuning parameter: Increasing the α penalizes the coefficients and thus shrinks them.

$$SS_{residuals} = \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

Model Prediction
↓
Observed Result

Ridge regression
(or "L2 regularization") minimizes:

$$RSS + \alpha \sum_{j=1}^p \beta_j^2$$

$$RSS + \alpha \sum_{j=1}^p |\beta_j|$$

Lasso regression
(or "L1 regularization") minimizes: